Three Essays in the Political Economy of Information

Elisa MOUGIN

Thesis supervised by
Julia CAGÉ, Associate Professor of Economics, Sciences Po Paris

defended on October, 1$^{\text{st}}$, 2021

Jury:

Pierre BOYER	Professeur, École Polytechnique – CREST (*referee*)
Julia CAGÉ	Associate Professor, Sciences Po Paris (*supervisor*)
Emeric HENRY	Associate Professor, Sciences Po Paris
Caroline LE PENNEC	Assistant Professor, HEC Montréal
Nicola MASTROROCCO	Assistant Professor, Trinity College Dublin
Maria PETROVA	ICREA Research Professor, Universitat Pompeu Fabra (*referee*)

Contents

Remerciements

Mes premiers remerciements vont tout naturellement à Julia Cagé, qui a supervisé cette thèse et m'a accompagnée tout au long de mon parcours de jeune chercheuse. Julia m'a fait confiance dès la première fois où nous nous sommes rencontrées en 2015 lorsque j'ai timidement passé la porte de son bureau pour lui parler de mon intérêt pour l'économie des médias et pour ses travaux en particulier. Qui m'aurait dit que débutait alors un chemin de six années qui m'a mené jusqu'à la rédaction de cette thèse. Quelle chance j'ai eue de pouvoir travailler avec toi ! Ton enthousiasme et ta curiosité m'ont portée d'un projet à l'autre, et je suis heureuse de la relation professionnelle et personnelle que nous avons tissée. J'ai appris de ton expertise, et ce, en particulier lors de la co-rédaction d'un chapitre de cette thèse, mais aussi de tes engagements et de ta conception du métier de chercheuse.

Ce qui m'amène à remercier mes co-auteurs.trices, indispensables à l'aboutissement des ces travaux. Le premier chapitre a été écrit avec Charles Louis-Sidois, aux côtés de qui j'ai découvert les méandres et les codes du monde académique. De longues heures de travail minutieux, nourries de beaucoup de rires et d'auto-dérision. Un grand merci à Caroline Le Pennec, pour la co-écriture du deuxième chapitre. J'ai beaucoup appris de ta rigueur, de tes capacités d'analyse, ainsi que de ton enthousiasme. Je suis très heureuse de te compter parmi les membres du jury.

Je tiens à remercier tout particulièrement Emeric Henry pour sa présence dans ce jury, ses conseils avisés et son accompagnement constructif, de Paris à Barcelone, en particulier pour le premier chapitre.

A very special thanks goes to Maria Petrova for having welcomed me during my visiting at the University Pompeu Fabra, and for giving me the opportunity to extend my stay in Barcelona in the fourth year of my doctoral studies. My work greatly improved thanks to your accurate questions and suggestions, and it means a lot to me to having you in the jury. I would like to warmly thank the other members of the jury, Pierre Boyer and Nicola Mastrorocco. Your respective works have been insightful and inspiring all along my own research, and I'm honored that you accepted to assess my work.

Je tiens également à remercier Roberto Galbiati, qui a accepté de co-encadrer cette thèse à ses débuts, et Benjamin Marx qui a rejoint mon comité de thèse à Sciences Po, pour son suivi toujours constructif.

J'ai rejoint en juillet dernier le Département des Etudes, des Affaires Economiques et de la Prospective au *Conseil Supérieur de l'Audioviduel*. Je remercie Christophe Cousin et Sébastien Lécou de leur confiance, ainsi que mes collègues qui m'ont tous chaleureusement accueillie. Travailler aujourd'hui au sein de cette équipe donne tout son sens à ces années d'études.

Ces remerciements sont très académiques, la vie est pourtant ailleurs et cette thèse doit tout autant à toutes les personnes qui m'ont aidée à garder les pieds sur terre. Mes amis d'enfance et camarades de fous rires, Malou, Mao et Yohann. Ensuite, "les dijonnais" : il y a maintenant plus de dix ans, nous entrions en prépa BL et je suis fière de ce que nous sommes devenus. Un grand merci à vous : Guillaume, Elia, Emilien, Maxime, Thomas, Léo, Rémi, et Victor, et bien entendu à Cyprien. Je remercie également Quentin et Andy, ainsi que la bande du Kenya.

Enfin, derrière ce travail se trouvent mes proches de chaque instant. Une page se tourne avec cette soutenance et c'est un nouveau chapitre qui commence. Je dédie cette thèse à ma filleule Joanne et à ses merveilleux parents, Morgane et Anatole, mes amis de toujours, dont le soutien et l'affection me sont si précieux.
Je remercie Florent, mon confident et fabuleux ami, Fatine et Yasmine, à mes côtés depuis le master à Sciences Po. Votre bienveillance à chaque instant rend notre amitié si chère. Chaque ligne de cette thèse a été écrite entre deux messages ou appels avec Salomé. Ceci à toute heure du jour et de la nuit, car il est rare que nous ayons été ensemble sur le même continent après avoir tant partagé à Dijon et à Paris. Merci d'avoir toujours été à mes côtés.

J'ai la chance d'avoir une famille nombreuse et unie : il serait trop long d'adresser un remerciement à chacun, mais je vous dois à tous, notamment cousins et cousines, d'avoir nourri ce lien à mes racines. Vous m'avez toujours rappelé l'importance des moments ensemble, conviviaux, festifs, toujours follement animés. Je dois beaucoup à la présence de Camille que j'embrasse ainsi que mes grands-parents Marie-Claire et Hugues. Un immense merci à mes petits frères Paul et Thomas, qui sont en réalité plus grands que moi par beaucoup d'aspects : leur confiance à toute épreuve, leur curiosité, nos débats enflammés ont été une source d'inspiration constante.
Enfin, je n'aurais pu avancer sans la présence, la gentillesse, et le soutien inconditionnel de mes parents, Christelle et Thierry. Merci à tous les deux de m'avoir toujours rappelé que les doutes, les peines et les joies n'ont de sens que si on les vit ensemble.

A tous, donc, ces sincères remerciements. Je suis heureuse et fière de vous avoir aujourd'hui à mes côtés. J'ai hâte de voir ce que demain nous réserve. Vivement la suite !

Introduction

An economic approach of information

The information market

The dissertation primarily studies the economics of the production of information through an analysis of its determinants and transformations. Through an economic perspective, the information market can be seen as shaped by many factors. For instance, in the case of the news, we can consider information as a good produced by editors or newsrooms to answer the demand coming for a variety of news consumers (Mullainathan and Shleifer, 2005). The characteristics of the market, such as the state of competition (Chan and Suen, 2008; Gentzkow and Shapiro, 2008; Cagé, 2019), the size of the market and media concentration (Noam, 2016; Prat, 2018), revenues and expected returns, or links to other economic sectors (Petrova, 2012; DellaVigna and Hermle, 2017; Durante et al., 2020) have been studied extensively as important inputs in the production function of newsrooms or content producers. While we can consider a market for news that can be analyzed and comprehended through the standard economic approach, it is commonly admitted that access to information directly and indirectly influences the organization and functioning of our societies (Cagé, 2015). Information can not be considered as a normal good, or at least not without considering it exempt from being at the origin of a wide range of externalities (Prat and Strömberg, 2011; Germano and Meier, 2013). The news content can be approached quantitatively (number of articles or news items, length, investments into investigative journalism) or qualitatively (inner quality of the information, used vocabulary, level of expertise of the speakers). The messages as outcomes provided to an audience do not only influence individuals in their own views and preferences, but also affect the way people collectively interact and shape social relationships and institutions. Gehlbach (2016) discusses the interaction between economic factors such as competition or concentration and media freedom and highlights the complexity of the nexus of market dynamics and the quality of the institutions. Indeed, the outcomes that can be affected by access to information are plentiful and span over all dimensions of

collective life. A large strand of the literature has studied how information affects individual behavior (see DellaVigna and Ferrara 2015 for a review of the effects of exposure to the media). Among many example, we can cite Jensen and Oster (2009) who show that access to cable TV in rural India significantly decreased the acceptability of violence against women and influenced more generally attitudes and values.[1] Another example is Bursztyn and Cantoni (2016) who highlight how exposure to Western programs – through geographic proximity to the boundary in East Germany – influenced consumption pattern among German citizens.

Discourse, media content, and political outcomes

More closely related to this thesis, access to information is at the core of our political systems and directly feed in how these systems should work. It had been shown that access to information can affect voters preferences and electoral competition (Piolatto and Schuett, 2015; Drago et al., 2013) , the state of the debate (Larcinese et al., 2011), and public policies or the setting of the political agenda (Stromberg, 2004; Djourelova and Durante, 2019). Community belongings are also influenced by exposure to the news and the standpoint through which the topics are considered (Depetris-Chauvin et al., 2020; Casey, 2015). Next, my the dissertation speaks to a larger literature that investigate the effect of the media on collective action and conflict (Della Vigna et al., 2010; Yanagizawa-Drott, 2014; Armand et al., 2020).

The scope of the analysis on content production goes beyond the media sector, and holds for other types of discourses that aim at reaching a more a less general audience. Hence, this discussion also encompasses political communication – shown to be a key factor in electoral matters (see for instanceLassen 2005; León 2017) – and access to information matters in a larger sense, for instance in candidate electoral campaigns' (Gerber et al., 2011; Pons, 2018; Le Pennec, 2020). Political connections to the media also make a difference, as studied for instance by Chiang and Knight (2011) in the case of endorsement of political candidates by newspapers.

Studying news coverage implies paying attention to the words or vocabulary – Djourelova (2020) shows that the ban of the word "illegal migrant" by the Associated Press news wire significantly impacted readers' view about immigration – or to the subjects favored by media outlets. For instance, Mastrorocco and Minale (2018) show that the decrease in the audience of traditional channels covering security issues leads to a decrease in concerns about crime in Italy. Another example is Beattie (2020) who emphasizes that advertising the car industry

---

[1]One can also refer to La Ferrara (2016) who document the multiple effects on entertainment TV on a large set of outcomes, such as education or health, in developing countries.

in the media shifts the tone of coverage of climate change issues. Editorial choices also include the level at which information is considered. A large literature has studied possible tradeoffs between national news and local news, and the provision of knowledge about local, national, or international issues interacts with all the above mentioned parameters (George and Waldfogel, 2006; Oberholzer-Gee and Waldfogel, 2009; Gao et al., 2020). Cagé (2020) also emphasizes that exacerbated local competition in the newspaper market can translate into fewer articles and reduce investigative capacity of the outlets in France. The recent work by Mastrorocco and Ornaghi (2020) similarly highlights that the degree of coverage of local news also depends on the structure of media ownership.[2] On top of the many factors previously discussed, the role played by media ownership patterns on coverage choices have been largely studied (Djankov et al., 2003). Numerous works have stressed that changes in media ownership can directly translate into changes of covered topics or tone of coverage (see for instance (Anderson and McLaren, 2012; Cagé et al., 2021).[3]

Media bias

All in all, choices of content are the result of the aggregation of many parameters and translate into a message, that is always one equilibrium, for instance the balance reached by a media owner and the journalists, that proposes one view of the world at a given time. For one story printed, one speech delivered, many others discourses, perspectives or viewpoints are not chosen. Then, this dissertation naturally speaks to the literature on media bias. Recall that the media sector is a two-sided market, and that on top of the many production parameters, the characteristics of the audience also enter the production function of the outlets that maximize their profits considering the demand coming from consumers(Mullainathan and Shleifer, 2005). Media bias can stem from the demand side of the market, since outlets strategies depend on people consumption habits and preferences (George and Waldfogel, 2003; Kennedy and Prat, 2019). I follow here Gentzkow and Shapiro (2006a) who define media bias as choices to slant information through "selective omission, choice of words and varying credibility ascribed to the primary source". In their work, they insist on the need for the source to appear as credible: then, the priors and beliefs of the consumers are key parameters and the authors stress that information is more often distorted to fit with ex-

---

[2]The authors study how the acquisitions of local TV stations by the Sinclair Group affected the coverage of crime issues in the media and document a decrease in the news coverage of local crime.

[3]Closely related to the question of media ownership is the link between government owned outlets and media slant or capture stemming from connections to the State. On this topic, see for instance, Enikolopov et al. (2011); Gehlbach and Sonin (2014); Knight and Tribin (2019) and Qin et al. (2018). Dependence on state advertising or subsidies has also been shown as a source of media bias (see for instance Di Tella and Franceschelli 2011).

pected priors of the audience.[4] Many studies have developed an economic approach of media bias, and investigate how bias affects citizens, such as Groseclose and Milyo (2005); Duggan and Martinelli (2011), or Gentzkow and Shapiro (2010) who document that readers in the United States tend to prefer newspapers sharing like-minded views.[5]

The past and the future of the news

Today, the media landscape is changing tremendously, notably with the entry of digital players in the media sectors and the rise of the social networks. Yet, one should recall that previous transformations, such as technological progress and regulation changes, have shaped the media landscape we have today. Indeed, the communication sector is in perpetual transformation, notably through changes in communication technology and the associated costs. Cotterlaz and Fize (2019) emphasize the effects of telegraph openings on the development of news agencies. The dynamics of the advertising markets is also crucial in the media sector (Anderson, 2018); Petrova (2011) sheds light on how the development of advertising markets helped strengthening a sustainable and free press in the United States, more independent from political pressures (see also Gentzkow et al. 2006 on the role of advertising on the rise of the fourth estate). It is also important to note that these economic factors, such as advertising revenues or sales, interact with each other, and that the emergence of new media participate in disrupting established business models. The example of the introduction of the radio has been documented for instance by Stromberg (2004) who highlights the effect of its expansion on public spending during the New Deal. Also, a large literature studies how the media landscape reacted to the introduction of TV (see among others Gentzkow and Shapiro 2006; Ellingsen and Hernæs 2018 or Angelucci et al. (2020)). The work by Angelucci and Cagé (2019) emphasizes how other outlets' revenues can be affected by the introduction of such a new technology: in the case of France, the introduction of advertising on TV created a significant drop in the revenues of national daily newspapers.

Moreover, as markets get larger with a more diverse supply, one can argue that it contributes to the maximization of the consumer welfare through a Pareto improvement the satisfaction of a plurality of consumer tastes (Waldfogel, 2017). However, the diversification on the supply side is often perceived as a source of broader polarization that would participate in the and emergence and the perpetuation of echo chambers (Putnam, 2000; Prior,

---

[4]This literature notably builds on the works on Bayesian persuasion mechanisms, such as Kamenica and Gentzkow (2011); Gentzkow and Shapiro (2010); Gitmez and Molavi (2018).

[5]It is also important to note here that the literature on bias and rhetorical choices is not restricted to the media but also pertains to the study of ideological slant, among politicians or policymakers for instance; this thesis speaks to a wider literature on the determinants of political platforms (see for instance Bonica (2014); Cruz et al. (2018).

2007). For instance, the diversification of television channels participated in the increasing share of audience dedicated to entertainment (Durante et al., 2019) and may also facilitate the development of more politically biased outlets. The effects of the right-wing leaning of the American channel *Fox News* on electoral outcomes have been documented by DellaVigna and Kaplan (2007) or Martin and Yurukoglu (2017). The debate around the political leaning of media outlets is also getting more complex with the digitization of the media sector. With the new technology and the entry of new digital players, traditional business models are disrupted and the development of free content online, notably information, gradually reshapes consumers' willingness to pay for content quality. Some studies have explored examples of such changes (see for instance Chiou and Tucker 2013 on the effects of the introduction of paywalls on readers' willingness to pay). Importantly, Cagé et al. (2020) document media's incentives to produce original content in an environment where it is easy to copy existing content. The debate around the incentives to invest in original content and quality journalism also echoes the discussion of the effects of the social networks on information production and citizens' attitudes toward the media.[6] More generally, today's debate around people's exposure to false information – explored for instance by Allcott and Gentzkow (2017) or Angelucci et al. (2020) in the United States, or by Barrera et al. (2020) in the case of France – and the reliability of news producers have a lot to do with the production of information in a digital world. The competition between traditional and new players such as the platforms and individual online activity is then at the core of today's redefinition of our way to produce and consume content (Cagé et al., 2020).

This thesis

The aim of the thesis is to delve into some of the determinants of the supply side of the information market, and consider the possible implications that those determinants can have on societies and political systems. The three chapters are independent from each other and can be read separately. They are three studies on three questions in media economics and aim at contributing to the debate on how information is produced and with with consequences.

As the main methodology, I use tools from applied microeconomics to quantitatively isolate relationships between one determinant of the production function of information and discourse, or content, as my outcome of interest. I also borrow to the political science literature and studies in communication to better discuss the pattern emphasized through data analysis.

---

[6]Also related is the literature on the Internet and government accountability (Enikolopov et al. 2018; Acemoglu et al. 2018; Zhuravskaya et al. 2020).

In the first chapter, I consider parameters that affect media capture, and how the perceptions of the likelihood to find a story to disclose and the characteristics of the market influence the different forms of pressures over media outlets over the world. In the second chapter, I look into the influence of money on political discourse and investigates how receiving donations from firms can affect candidates discourse, in an analysis of political manifestos issued by candidates to the French legislative election. In the third chapter, I study the impact of technological change on media content and political behavior. More precisely, I look at the effect of the introduction of digital TV in Kenya on news content and on political preferences during the presidential election of 2017.

Hence, the three chapters of this dissertation illustrate the complexity of the media landscape and of the determinants of the political discourse. Building on the existing results from the literature and using a large range of methods, I try to contribute to the general debate on the challenges pertaining to today's world of information and entertainment, in its multifaceted aspects.

Chapter 1: Silence the Media or the Story

This chapter is coauthored with Charles Louis-Sidois.

The first chapter of this thesis focuses on media capture and its determinants. By media capture, I refer to the approach developed by Stiglitz (2017) who defines it as "an economics term that describes what happens when regulators become overly empathetic or supportive of those they are meant to be regulating". The author delves into the different forms it can take; the more obvious being media capture via ownership, but it can also happen through financial incentives, or pressures such as of censorship. These various forms have been extensively studied in the literature. For instance, Dyck et al. (2008); Durante and Knight (2012) examine capture through ownership; McMillan and Zoido (2004) study direct payments to journalists; and Gratton (2015) focuses on the impact of threats on journalistic activity. The goal of the chapter of this dissertation is to address the complexity of the phenomenon by studying together the different types of capture. To better understand what environments explain the prevalence of some pressures against the media, it combines theoretical and empirical methods.

On top of the various determinants that affect coverage choices, the editorial agenda can also be determined by the likelihood of the newsrooms to find out some news to provide to the audience. Hence, prohibitive costs of investigation, or limited size of the audience, can deter journalists to launch an investigation. We consider how these parameters enter into

the choice of an agent who may be interested in erasing a story, and introduce the timing of capture as a key parameter of the preferred strategy.

The key distinction we draw between the different strategies of media capture is the control of the agenda of journalistic investigation, which is contingent on the timing of capture. On one side of the spectrum, media ownership gives the principal – who can be a lobby, a politician, or any agent willing to interfere with the journalistic work – a total control of the agenda. This strategy is preventive as it implies an ex-ante acquisition of the media. We call this strategy internal capture. Instead, the principal can refrain from setting the agenda and propose an ex-post payment after the investigation process, typically a bribe, in exchange for the suppression of negative news stories. We refer to this reactive option as external capture.

First, we build on the seminal work by Besley and Prat (2006) and propose a model in which a principal is corrupt. To obtain a signal about the type of the principal, media outlets strategically choose how much to investigate. Receiving and publishing a signal indicating that the principal is corrupt provides a positive payoff to a media outlet, but negatively affects the principal. The principal has two strategies to prevent the disclosure of the signal. First, it can make a transfer to a media outlet at the beginning of the game and prevent it from investigation (internal capture). Alternatively, the principal can make an offer to a media outlet which has investigated and has received a signal, in exchange for its suppression (external capture).

The key predictions of our model relate to the effect of perceived corruption on media capture. Firstly, we expect internal capture to increase with perceived corruption. An early agreement is preferred by the principal if media outlets would exert high investigation efforts, which is the case if they initially believe that the principal is corrupt. Turning to external capture, we expect a positive effect of perceived corruption if the media market is competitive, but a negative effect if it is less developed. This prediction is driven by two opposite forces: an increase in perceived corruption results in more investigation from media outlets, which in turn increases the necessity of external capture. On the other hand, perceived corruption increases internal capture, which renders external capture unnecessary. When the media market is small, the second effect dominates, which explains our result on external capture.

In the second part of the chapter, we bring these predictions to the data and provide empirical support to the theoretical findings. The positive relationship between media capture and corruption is well established, but the causal relation is hard to identify (Prat and Strömberg, 2013). Our idea here is to use an exogenous shock on newsrooms' likelihood to investigate, which is independent from the specifics of the media market in the short run.

To do so, we use the revelations of the Panama Papers and the large worldwide coverage it received.[7] We exploit the fact that countries were differently exposed to the shock, depending on the number of offshore entities mentioned in the leaks, and look at cross-country variations in the adjustments of capture strategies.

We implement a difference-in-differences strategy and first show that the revelations did affect the perceptions of corruption within countries. We show that in countries more exposed to the scandals, perceived corruption increased by 1.2 percentage point in the aftermath of the revelation. We build on this result to discuss how the change in perceptions could have affected the media landscape and capture strategy. To do so, we use data provided by *Reporters Without Borders* for 187 countries from 2012 to 2018. For internal capture, we consider conflicts of interest with media owners and the share of owners who have interests in other economic sectors. External capture is proxied by outside payments and in-kind benefits to journalists.

Our empirical results are in line with our theoretical predictions. First, we document a 9-percentage points increase in the share of owners with interests in other economic sectors, which supports the prediction of a positive relationship between perceptions of corruption and internal capture. Second, we provide evidence that the increase in perceived corruption translates into a lower prevalence of external capture in countries where the media landscape is the least developed, while we find the opposite result in countries where the media landscape is larger.

The chapter is an attempt to better understand the extent of media capture today and to propose an approach to address its complexity. It is based on perceived corruption as a key determinant of media entry, but other parameters, such as media competition, also affect the principal's tradeoff and could impact capture strategies. Our model and empirical findings are a first step in understanding that the forms of capture are crucial to protect the freedom of the press worldwide.

Chapter 2: Money and Ideology

This chapter is coauthored with Julia Cagé and Caroline Le Pennec.

The second chapter aims at studying the impact of corporate donations on the content of candidates' campaign communication. More precisely, we focus on the French legislative elections in the 1990's, and examine the impact of donations from legal entities on candidates'

---

[7]The Panama papers consisted of a leak of documents that explicitly mentioned individuals or firms involved in tax avoidance schemes, released in 2016 by the International Consortium of Investigative Journalists.

choices of rhetoric in their campaign manifestos. The impact of money in politics have been widely studied, notably the impact of campaign spending on electoral success or activity of elected representatives (see for instance Levitt 1994; Ansolabehere et al. 2003; Kalla and Broockman 2018 or Bekkouche et al. 2020); however, little is known on how the type of revenues affects candidates' rhetoric at the campaign stage. Yet, we know thanks to Feltovich and Giovannoni (2015); Kendall et al. (2015) or Cruz et al. (2018) among others, that campaign messages are an important determinant of voters' choices.

We construct a novel dataset that combines information on donations received by candidates running for parliamentary seats in 555 constituencies, and the individual campaign manifestos they issued before the elections. Our dataset contains 10,285 single manifestos, from Le Pennec (2020) and our own data collection from the National Archives. Moreover, our dataset includes detailed information on campaign revenues and the amounts given by firms to each candidate, on top of details on their individual characteristics (party, gender, other political mandates, ...).

We exploit an historical shock on corporate donations and use a difference-in-differences approach to estimate the causal impact of donations on the content of campaign communication. In 1995, a reform banned corporate donations to political candidates: this law marked an important change in party and election financing. Since then, only "natural" persons (i.e. individuals) have been allowed to make political donations.[8]

The main goal of the study is to estimate whether the ban had an impact on candidate discourse, and if so, along what dimensions the rhetoric could have been affected. We formulate four hypotheses to better comprehend how donations from firms could influence the content of the manifestos. First, corporate donations could push candidates to run a more local campaign – e.g. focused on their district – or a more national campaign – e.g. focused on party organizations and prominent politicians. Second, candidates could use more polarized or more moderate language in response to donations. Third, corporate donations could play a role on the prevalence of some policy topics – such as the economy or homeland security – over others. Finally, they can influence the quality of a candidate's communication, which we proxy using the level of originality of the manifestos. To convert these hypotheses into quantitative indices that can be quantitatively estimated, we apply various methods of computational text analysis – borrowed from Gentzkow et al. (2019) and Le Pennec (2020). We construct four distinct measures that represent the dimensions of language that individual politicians have control over and may adjust in response to the

---

[8]Since 1988, the "Commission Nationale des Comptes de Campagne et des Financements Politiques" is in charge of checking and approving the accounts of candidates' campaigns. Moreover, we document that the 1995 was unanticipated by the candidates.

campaign contributions they receive.

Because these campaign contributions are not randomly allocated across candidates, estimating their causal effect requires an identification strategy that handles their endogenous nature. In 1993, about a third of all candidates running for French legislative elections received at least one donation from a corporate entity. On average, men tend to receive more corporate donations than women, while incumbents and re-runners receive more donations than first-time candidates. We also find that candidates from mainstream parties receive more – and larger – donations, than "niche" and marginal parties' candidates. We exploit the fact that candidates varied significantly in their reliance on such donations before the ban and estimate a difference-in-differences model to quantify the impact of receiving corporate donations in 1993.

We then estimate our difference-in-differences model using as outcomes the four indices previously described. First, with regards to the local index, we find an economically meaningful impact which suggests that receiving more donations encourages candidates to advertise their local presence in their campaign communication. According to our estimates, a one-standard-deviation increase in corporate donations increases the local index by 16% of a standard deviation. The effect is particularly high for parties that are newer on the political scene (green and far right candidates). Second, we find no significant impact of receiving donations on either the left-right score of candidate manifestos, or their extremeness. However, when looking at possible heterogeneity across parties, we find that extreme candidates do adjust their speech. Indeed, while the effect on the left-right score remains small and mostly insignificant for the three historically dominant parties (right-wing, Socialist and Communist), Green candidates move to the right of the language scale and far-right candidates move to the left. Third, we consider policy topics in the manifestos by defining four main categories of interest: homeland and administration, foreign policy, economy, and social issues. We find evidence that corporate donations encourage candidates to devote more space in their campaign communication to economic issues, at the expense of foreign policy and social issues. For instance, a one-standard-deviation increase in corporate donations raises the probability of focusing on economic issues by 1.6 percentage points. Last, we find a significant and positive impact of corporate donations on the originality index, suggesting that more effort are put by candidates to issue a personalized manifesto as compared to other candidates running for the same party.

The last part of the chapter aims at rationalizing these empirical findings. We argue that donations have an "electoral effect" on candidates: receiving contributions from local corporate donors raises the salience of local and economic topics. As their perception of voters' most important issues changes, candidates adjust their persuasive campaign com-

munication accordingly. This effect is particularly significant for marginal candidates and outsiders, whose prior beliefs regarding voters' preferences may not be as precise as mainstream candidates'.

This interpretation is our preferred one, and we empirically support this claim by considering other types of possible mechanisms. For instance, we could have considered that corporate donations may have resulted in more local references in the electoral discourse because of a "resource effect". An increased campaign revenue, no matter its sources, may enable politicians to run a better-quality campaign with a more targeted communication, better tailored to their local electorate. However, a more in-depth analysis brings to light substantial heterogeneity across donors. Indeed, we find that the effect of donations on the local index is driven by small donors, that is to say firms who gave to only one candidate in one district. This heterogeneity result seems to rule out this first mechanism as the latter would have pertained to all type of donors.

Another possible explanation is that corporate donors may expect particularized benefits in exchange for their (service-induced) donations. This "quid-pro-quo effect" interpretation is consistent with the fact that local donors, with primarily local interests, push candidates to put forward their local presence in their manifesto – and advertise the influence they can exert at the local level to serve their donors' interests. Yet, we do not find any significant impact of corporate donations on either the quantity or the content of questions to the government and debate interventions among elected representatives. From these results, we cannot conclude that politicians favor their corporate donors once in power. Then, while we cannot rule out the existence of some quid-pro-quo relationship between corporate donors and politicians, such a mechanism is unlikely to be the main channel of the candidates' rhetorical adjustment to corporate donations.

Overall, our results suggest that connections with private corporations may alter politicians' behavior even in the absence of corruption and quid-pro-quo agreements. We provide evidence that campaign contributions may influence citizens through their indirect effect on politicians: by affecting candidates' perceptions of voters' concerns, contributions shape the content of campaign messages distributed during the electoral season and, ultimately, the information voters use to form their decision.

Chapter 3: TV in times of political uncertainty

In the third chapter of this thesis, I study an example of a technological change in the media landscape – namely the introduction of digital television in Kenya – to document the impact of the introduction of a new medium on voting patterns. In Kenya, we observe a

dramatic increase in signal availability of television in the last decade, which gives a very interesting empirical framework to study the fallouts of the entry of new technology in a sub-Saharan African country. Between 2013 and 2017, more than 2,500 sublocations benefited from the migration from analog to digital television, representing more than 30% of the population who got access to the television signal. This increase in TV penetration is largely due to the installation of television transmitters, in the context of the digital migration plan in Kenya and the adoption of *Digital Video Broadcasting - Terrestrial* transmission standards (DVB-T and DVB-T2), financed and supervised by the government with the support of the *International Telecommunication Union.*

This chapter builds on the literature on the impact of TV on voting outcomes (Gentzkow and Shapiro, 2006; Ellingsen and Hernæs, 2018; Durante et al., 2019) and discusses its relevance in a developing world setting. Moreover, I exploit the very peculiar context of the 2017 Kenyan election to examine deeper the possible heterogeneous effects of exposure to TV in periods of political uncertainty. The nullification of the 2017 election by the Supreme Court completely changed the political odds in a couple of weeks only. Indeed, the decision of the Kenyan Supreme Court to nullify the election was announced on September 1st 2017 after the incumbent President had been declared the winner on August 2017. The cancellation followed a complaint lodged by the opposition party and was based on the misalignment in the reporting of the results to the electoral commission between the different polling centers. The statement marked a breakthrough in Kenyan politics, and the debate around the nullification and the organization of new polls went together with a wave of protests conducted by the opposition (Wairuri, 2017). The organization of a repeat ballot, that took place three months after the first election, and that largely boycotted by the opposition, dramatically raised the level of uncertainty. Therefore, I compare the effect of TV on turnout in the first election – that we can consider as a "normal" context ballot – with the repeat election, and discuss how the climate of political crisis may be an important factor in the heterogeneity of influence of media exposure on voters.

Before estimating the impact of TV on turnout, I investigate the main determinants of TV reception using topographic data and socio-economic controls. Indeed, the installation of TV transmitters and exposure to the broadcast is very likely to be endogenous. It is plausible that the transmitters are installed in strategic places, to reach larger or richer audience, and that the new receivers have voting attitudes that are not representative of the whole population. To measure the real influence of exposure to broadcasts, we need to rely on the assumption that locations with or without access to TV in 2017 are similar in unobserved characteristics, conditional on observables. I construct a geo-coded dataset of Kenyan polling centers and TV transmitters and compute the fourth polynomial of the

distance between each voting center and the nearest transmitter. The dataset also includes terrain ruggedness measures, distances to the capital or urban centers, urbanization status as well as demographic and economic covariates, measured at a fine-grained level. In line with the existing literature (Olken, 2009; Yanagizawa-Drott, 2014; Miner, 2015), I find that the distance to the transmitter and terrain features are the main predictors of signal reception. For instance, a 1-km increase in the distance to the transmitter increases the probability to receive the signal by around 60%.

I conduct a series of exogeneity checks to test for the validity of the identification assumption, i.e. that signal availability isn't significantly correlated with covariates that also influence political participation, conditional on observables. I find that conditional on signal predictors, political preferences before 2013, or ethnic belongings, are not significantly correlated with TV reception in 2017. I run similar checks on economic and social covariates – before and at the time of the election – and do not find evidence of bias in the allocation of TV, conditional on the propagation determinants.

Thereafter, I estimate a difference-in-differences model to measure the impact of access to TV on turnout for both the first and the recall election, where the treated polling stations are those that were granted access to the signal between 2013 and 2017, controlling for the signal predictors previously identified. I find that exposure to the broadcasts didn't influence turnout in the first *regular* election. However, it significantly impacted turnout in the repeat election and this effect is heterogeneous across political camps, that I define depending on whether a polling station had voted in favor of the incumbent in the previous election of 2013. In the repeat election, turnout dropped in pro-opposition polling stations by 4 percentage points, but it is 7 percent higher in polling centers that voted in favor of the incumbent in 2013. Another specification of the model suggests that a 1-percent increase in the score of the incumbent adds around 0.15 percentage point to the baseline effect of TV. In other words, exposure to TV seems to have deterred voters to cast a new ballot where the opposition ranked first, while it rather had amplified support to the incumbent where the latter was leading the race.

The last part of the chapter aims at uncovering the possible channels of TV influence to better understand why the electoral context plays such a role on TV influence. I first study the content of the TV broadcasts and analyze a dataset made of 86,189 news stories broadcast on TV. Using text analysis methods, I document the type of news stories that were on air at the time of the election. I provide evidence of the large emphasis put on the climate of tension and on the ongoing riots in the country by the TV channels in the aftermath of the nullification of the first election. For instance, in the two weeks before the repeat election, the share of conflict-related news on TV increased by 10%. Moreover, I

emphasize a drop in the diversity of reported news stories, that indicates a shift from the usual coverage choices, and a large focus on the political crisis.

Also, the strength of TV as a medium can be explained by collective exposure in public places. In Kenya, around 40% of the population owns a TV set today, while more than a half of the population gets informed through TV broadcasts at least once a month. This gap between TV ownership and viewership is suggestive of the collective dimension in exposure to TV which is often watched in public places. I provide evidence that attendance of cafes or restaurants is higher in places covered by TV, and suggest that this collective dimension of news watching can translate into voting coordination mechanisms.

More generally, this chapter relates to a larger literature on technological changes and impact of exposure to the media. It contributes to the growing debate on the transformations in access to information in developing countries. It suggests that individual reactions to uncertainty can be amplified by the media, notably through shifts in editorial choices, which can end up affecting political behavior.

References

Acemoglu, D., Hassan, T. A., and Tahoun, A. (2018). The power of the street: Evidence from egypt's arab spring. *The Review of Financial Studies*, 31(1):1–42.

Allcott, H. and Gentzkow, M. (2017). Social media and fake news in the 2016 election.

Anderson, S. (2018). Legal Origins and Female HIV. *American Economic Review*, 108(6):1407–1439.

Anderson, S. P. and McLaren, J. (2012). Media Mergers And Media Bias With Rational Consumers. *Journal of the European Economic Association*, 10(4):831–859.

Angelucci, C. and Cagé, J. (2019). Newspapers in times of low advertising revenues. *American Economic Journal: Microeconomics*, 11(3):319–64.

Angelucci, C., Cagé, J., and Sinkinson, M. (2020). Media competition and news diets. Technical report, National Bureau of Economic Research.

Ansolabehere, S., de Figueiredo, J. M., and Snyder, J. M. J. (2003). Why is There so Little Money in U.S. Politics? *Journal of Economic Perspectives*, 17(1):105–130.

Armand, A., Atwell, P., and Gomes, J. F. (2020). The reach of radio: Ending civil conflict through rebel demobilization. *American Economic Review*, 110(5):1395–1429.

Barrera, O., Guriev, S., Henry, E., and Zhuravskaya, E. (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of Public Economics*, 182:104123.

Beattie, G. (2020). Advertising and media capture: The case of climate change. *Journal of Public Economics*, 188:104219.

Bekkouche, Y., Cagé, J., and Dewitte, E. (2020). The Heterogeneous Price of a Vote: Evidence from Multiparty Systems, 1993-2017. CEPR Discussion Papers 15150, C.E.P.R. Discussion Papers.

Besley, T. and Prat, A. (2006). Handcuffs for the grabbing hand? media capture and government accountability. *American Economic Review*, 96(3):720–736.

Bonica, A. (2014). Mapping the Ideological Marketplace. *American Journal of Political Science*, 58(2):367–386.

Bursztyn, L. and Cantoni, D. (2016). A tear in the iron curtain: The impact of western television on consumption behavior. *Review of Economics and Statistics*, 98(1).

Cagé, J. (2015). The Economics of the African Media. In Monga, C. and Lin, J. Y., editors, *The Oxford Handbook of Africa and Economics: Policies and Practices*, pages 605–625. Oxford University Press.

Cagé, J. (2019). Media competition, information provision and political participation: Evidence from french local newspapers and elections, 1944–2014. *Journal of Public Economics*, page 104077.

Cagé, J. (2020). Media competition, information provision and political participation: Evidence from French local newspapers and elections, 1944–2014. *Journal of Public Economics*, 185.

Cagé, J., Hervé, N., and Mazoyer, B. (2020). Social media and newsroom production decisions. *SSRN Working Paper*.

Cagé, J., Hervé, N., and Viaud, M.-L. (2020). The Production of Information in an Online World. *The Review of Economic Studies*, 87(5):2126–2164.

Cagé, J., Hengel, M., Hervé, N., and Urvoy, C. (2021,). Hosting media bias: Evidence from the universe of french television and radio shows, 2002-2020.

Casey, K. (2015). Crossing party lines: The effects of information on redistributive politics. *American Economic Review*, 105(8):2410–2448.

Chan, J. and Suen, W. (2008). A spatial theory of news consumption and electoral competition. *Review of Economic Studies*, 75(3):699–728.

Chiang, C. F. and Knight, B. (2011). Media bias and influence: Evidence from newspaper endorsements. *Review of Economic Studies*, 78(3):795–820.

Chiou, L. and Tucker, C. (2013). Paywalls and the demand for news. *Information Economics and Policy*, 25(2):61–69.

Cotterlaz, P. and Fize, E. (2019). Information in the first globalization: News agencies and trade. *SSRN Working Paper*.

Cruz, C., Keefer, P., Labonne, J., and Trebbi, F. (2018). Making Policies Matter: Voter Responses to Campaign Promises. Working Paper 24785, National Bureau of Economic Research.

Della Vigna, S., Enikolopov, R., Mironova, V., Petrova, M., and Zhuravskaya, E. (2010). Cross-Border Effects of Foreign Media: Serbian Radio and Nationalism in Croatia.

DellaVigna, S. and Ferrara, E. L. (2015). Chapter 19 - Economic and Social Impacts of the Media. In Anderson, S. P., Waldfogel, J., and Strömberg, D., editors, *Handbook of Media Economics*, volume 1 of *Handbook of Media Economics*, pages 723–768. North-Holland.

DellaVigna, S. and Hermle, J. (2017). Does conflict of interest lead to biased coverage? evidence from movie reviews. *Review of Economic Studies*, 84(4):1510–1550.

DellaVigna, S. and Kaplan, E. (2007). The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234.

Depetris-Chauvin, E., Durante, R., and Campante, F. (2020). Building nations through shared experiences: Evidence from african football. *American Economic Review*, 110(5):1572–1602.

Di Tella, R. and Franceschelli, I. (2011). Government Advertising and Media Coverage of Corruption Scandals. *American Economic Journal: Applied Economics*, 3(4):119–151.

Djankov, S., McLeish, C., Nenova, T., and Shleifer, A. (2003). Who Owns the Media? *Journal of Law and Economics*, 46(2):341–381.

Djourelova, M. (2020). Media persuasion through slanted language: Evidence from the coverage of immigration.

Djourelova, M. and Durante, R. (2019). Media attention and strategic timing in politics: Evidence from us presidential executive orders.

Drago, F., Nannicini, T., and Sobbrio, F. (2013). Meet the Press: How Voters and Politicians Respond to Newspaper Entry and Exit. IZA Discussion Papers 7169, Institute for the Study of Labor (IZA).

Duggan, J. and Martinelli, C. (2011). A spatial theory of media slant and voter choice. *Review of Economic Studies*, 78(2):640–666.

Durante, R. and Knight, B. (2012). Partisan control, media bias, and viewer responses: Evidence from berlusconi's italy. *Journal of the European Economic Association*, 10(3):451–481.

Durante, R., Peydró, J.-L., et al. (2020). Media capture by banks.

Durante, R., Pinotti, P., and Tesei, A. (2019). The political legacy of entertainment tv. *American Economic Review*, 109(7):2497–2530.

Dyck, A., Volchkova, N., and Zingales, L. (2008). The corporate governance role of the media: Evidence from russia. *The Journal of Finance*, 63(3):1093–1135.

Ellingsen, S. and Hernæs, Ø. (2018). The impact of commercial television on turnout and public policy: Evidence from Norwegian local politics. *Journal of Public Economics*, 159.

Enikolopov, R., Petrova, M., and Sonin, K. (2018). Social media and corruption. *American Economic Journal: Applied Economics*, 10(1):150–74.

Enikolopov, R., Petrova, M., and Zhuravskaya, E. (2011). Media and Political Persuasion: Evidence from Russia. *American Economic Review*, 101(7):3253–85.

Feltovich, N. and Giovannoni, F. (2015). Selection vs. accountability: An experimental investigation of campaign promises in a moral-hazard environment. *Journal of Public Economics*, 126:39–51.

Gao, P., Lee, C., and Murphy, D. (2020). Financing dies in darkness? The impact of newspaper closures on public finance. *Journal of Financial Economics*, 135(2):445–467.

Gehlbach, S. (2016). Media freedom: What matters?

Gehlbach, S. and Sonin, K. (2014). Government Control of the Media. *Journal of Public Economics*, 118:163–171.

Gentzkow, M., Glaeser, E., and Goldin, C. (2006). The Rise of the Fourth Estate: How Newspapers Became Informative and Why It Mattered. In *Corruption and Reform: Lessons from America's Economic History*. National Bureau of Economic Research.

Gentzkow, M. and Shapiro, J. M. (2006a). Media Bias and Reputation. *Journal of Political Economy*, 114(2):280–316.

Gentzkow, M. and Shapiro, J. M. (2006b). Media Bias and Reputation. *Journal of Political Economy*, 114(2).

Gentzkow, M. and Shapiro, J. M. (2008). Competition and Truth in the Market for News. *Journal of Economic Perspectives*, 22(2):133–154.

Gentzkow, M. and Shapiro, J. M. (2010). What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica*, 78(1):35–71.

George, L. and Waldfogel, J. (2003). Who Affects Whom in Daily Newspaper Markets? *Journal of Political Economy*, 111(4):pp. 765–784.

George, L. M. and Waldfogel, J. (2006). The New York Times and the Market for Local Newspapers. *American Economic Review*, 96(1):435–447.

Gerber, A., Gimpel, J., Green, D., and Shaw, D. (2011). How Large and Long-lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a Randomized Field Experiment. *American Political Science Review*, 105(01):135–150.

Germano, F. and Meier, M. (2013). Concentration and self-censorship in commercial media. *Journal of Public Economics*, 97(C):117–130.

Gitmez, A. and Molavi, P. (2018). Media capture: a bayesian persuasion approach.

Gratton, G. (2015). The sound of silence: Political accountability and libel law. *European Journal of Political Economy*, 37:266–279.

Groseclose, T. and Milyo, J. (2005). A Measure of Media Bias. *Quarterly Journal of Economics*, 120(4):1191–1237.

Jensen, R. and Oster, E. (2009). The Power of TV: Cable Television and Women's Status in India. *The Quarterly Journal of Economics*, 124(3):pp. 1057–1094.

Kalla, J. L. and Broockman, D. E. (2018). The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments. *American Political Science Review*, 112(1):148–166.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Kendall, C., Nannicini, T., and Trebbi, F. (2015). How Do Voters Respond to Information? Evidence from a Randomized Campaign. *American Economic Review*, 105(1):322–353.

Kennedy, P. J. and Prat, A. (2019). Where Do People Get Their News? *Economic Policy*, 34(97):5–47.

Knight, B. and Tribin, A. (2019). Opposition media, state censorship, and political accountability: Evidence from chavez's venezuela.

La Ferrara, E. (2016). Mass Media and Social Change: Can We Use Television to Fight. *Journal of the European Economic Association*, 14(4):791–827.

Larcinese, V., Puglisi, R., and Snyder, J. M. (2011). Partisan bias in economic news: Evidence on the agenda-setting behavior of U.S. newspapers. *Journal of Public Economics*, 95(9-10):1178–1189.

Lassen, D. D. (2005). The Effect of Information on Voter Turnout: Evidence from a Natural Experiment. *American Journal of Political Science*, 49(1):pp. 103–118.

Le Pennec, C. (2020). Strategic campaign communication: Evidence from 30,000 candidate manifestos. SoDa Laboratories Working Paper Series 2020-05, Monash University, SoDa Laboratories.

León, G. (2017). Turnout, political preferences and information: Experimental evidence from peru. *Journal of Development Economics*, 127:56–71.

Levitt, S. D. (1994). Using Repeat Challengers to Estimate the Effect of Campaign Spending on Election Outcomes in the U.S. House. *Journal of Political Economy*, 102(4):777–798.

Martin, G. J. and Yurukoglu, A. (2017). Bias in Cable News: Persuasion and Polarization. *American Economic Review*, 107(9):2565–2599.

Mastrorocco, N. and Minale, L. (2018). News media and crime perceptions: Evidence from a natural experiment. *Journal of Public Economics*, 165:230–255.

Mastrorocco, N. and Ornaghi, A. (2020). Who watches the watchmen? local news and police behavior in the united states.

McMillan, J. and Zoido, P. (2004). How to Subvert Democracy: Montesinos in Peru. *Journal of Economic Perspectives*, 18(4):69–92.

Miner, L. (2015). The unintended consequences of internet diffusion: Evidence from Malaysia. *Journal of Public Economics*.

Mullainathan, S. and Shleifer, A. (2005). The Market for News. *American Economic Review*, 188(2):355–408.

Noam, E. M. (2016). *Who owns the world's media?: media concentration and ownership around the world*. Oxford University Press.

Oberholzer-Gee, F. and Waldfogel, J. (2009). Media Markets and Localism: Does Local News en Espanol Boost Hispanic Voter Turnout? *American Economic Review*, 99(5):2120–2128.

Olken, B. A. (2009). Corruption perceptions vs. corruption reality. *Journal of Public Economics*, 93(7-8):950–964.

Petrova, M. (2011). Newspapers and Parties: How Advertising Revenues Created an Independent Press. *The American Political Science Review*, 105(4):pp. 790–808.

Petrova, M. (2012). Mass media and special interest groups. *Journal of Economic Behavior and Organization*, 84(1):17–38.

Piolatto, A. and Schuett, F. (2015). Media competition and electoral politics. *Journal of Public Economics*, 130:80–93.

Pons, V. (2018). Will a five-minute discussion change your mind? A countrywide experiment on voter choice in France. *American Economic Review*, 108(6):1322–1363.

Prat, A. (2018). Media power. *Journal of Political Economy*, 126(4):1747–1783.

Prat, A. and Strömberg, D. (2011). The Political Economy of Mass Media. Discussion paper, CEPR.

Prat, A. and Strömberg, D. (2013). The political economy of mass media. *Advances in economics and econometrics*, 2:135.

Prior, M. (2007). *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*. Cambridge Studies in Public Opinion and Political Psychology. Cambridge University Press.

Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster, New York.

Qin, B., Strömberg, D., and Wu, Y. (2018). Media Bias in China. *American Economic Review*, 108(9):2442–2476.

Stiglitz, J. (2017). Toward a Taxonomy of Media Capture. In Schiffrin, A., editor, *In the Service of Power: Media Capture and the Threat to Democracy*, pages 9–17. Center for International Media Assistance.

Stromberg, D. (2004). Political Mass Media Competition, and Public Policy. *Review of Economic Studies*, 71(1):265–284.

Wairuri, K. (2017). Ridng the Crest of the Wave? The 2017 election and stagnation of Kenya's democratization process. *Note de l'Observatoire de l'Afrique de l'Est.*

Waldfogel, J. (2017). How digitization has created a golden age of music, movies, books, and television. *Journal of economic perspectives*, 31(3):195–214.

Yanagizawa-Drott, D. (2014). Propaganda and Conflict: Theory and Evidence From the Rwandan Genocide. *Quarterly Journal of Economics*, 129(4):1947–1994.

Zhuravskaya, E., Petrova, M., and Enikolopov, R. (2020). Political Effects of the Internet and Social Media. *Annual Review of Economics*, Forthcomin.

Chapter 1

Silence the Media or the Story: Theory and Evidence of Media Capture

This chapter is joint work with Charles Louis Sidois.

Abstract

We explore a theory of media capture where a principal can either influence journalistic investigation (internal capture) or let the media investigate and pay to suppress news stories at the publication stage (external capture). We predict that the likelihood of internal capture increases with perceived corruption. Instead, external capture decreases with perceived corruption if the media market is sufficiently competitive. We study a sample of 166 countries between 2012 and 2018, using new survey data collected by Reporters Without Borders and a set of corruption indicators. We use the revelation of the Panama Papers as a shock to perceived corruption. With a difference-in-differences identification strategy based on cross-country variation in exposure to the shock, we find support for our theoretical predictions.

1 Introduction

Media capture poses a well identified threat to the freedom of the press. It occurs when a principal – for instance the government, an interest group or a firm – influences the media in order to manipulate the news. Media capture can take multiple forms, such as media ownership, pressure through advertising, bribes and even jailing and physical violence against journalists (Durante and Knight 2012, DellaVigna and Hermle 2017). As shown in Figure 1.1, those strategies are widespread and, more importantly, their prevalence varies across regions. This paper focuses on this variation and shows that the forms of media capture depend on the environment. In particular, we document that an exogenous increase in the perception of corruption leads to a larger likelihood of capture through media ownership. Moreover, in countries where the media sector is not sufficiently developed, an increase in perceived corruption decreases the bribing of journalists.

The choice between the different forms of media capture is an understudied topic, which limits the efficiency of programs promoting the freedom of the press (Reporters Without Borders 2016). According to Yavuz Baydar, a Turkish journalist: *"While the world is focused on the issue of jailed journalists in Turkey, [...] the kiss of death to our profession has been bestowed by owners who consciously destroy editorial independence, fire journalists who voice skepticism and dissent and block investigative reporting"*[1]. Ensuring a well-functioning media landscape requires a comprehensive approach to media capture which accounts for its different materializations. Thus, we aim to examine, both theoretically and empirically, how the environment influences the choice of media capture.

The key distinction we draw between the different strategies of media capture is the control of the agenda of journalistic investigation, which is contingent on the timing of capture. On one side of the spectrum, media ownership gives the principal total control of the agenda. This strategy is preventive as it implies an ex-ante acquisition of the media. We call this strategy *internal capture*. Instead, the principal can refrain from setting the agenda and propose an ex-post payment after the investigation process, typically a bribe, in exchange for the suppression of negative news stories. We refer to this reactive option as *external capture*.

We propose a theoretical model explaining how the prevalence of internal and external capture varies with the perception of corruption. Drawing on the seminal work by Besley and Prat (2006), a principal who could be a politician, a firm or a lobby, is corrupt[2] with some probability. Media outlets strategically choose how much to investigate. Investigation

---

[1] *New York Times*, 19 July 2013.

[2] We stick to this terminology throughout the paper but the model is more general and covers all types of wrongdoing.
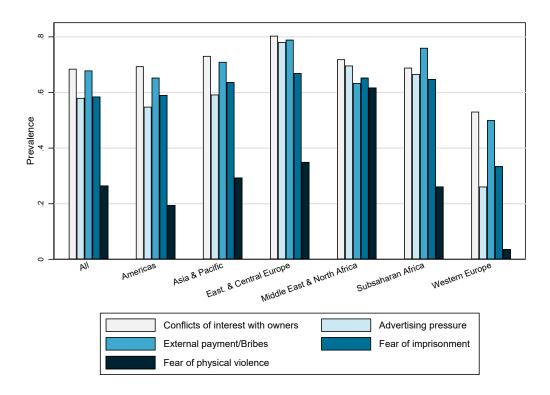
Figure 1.1: Types of media capture

is costly and positively affects the probability to obtain a perfectly informative signal about the type of the principal. For a media outlet, receiving a signal indicating that the principal is corrupt and publishing it provides a positive payoff, but negatively affects the principal. If several media outlets publish the signal simultaneously, they share the publication payoff. To prevent the disclosure of the signal, the principal can make two types of offers: internal capture takes place at the beginning of the game, before the investigation decisions, while external capture occurs between the realization of the signals and publication. Both are take-it-or-leave-it offers and, if a media outlet accepts, it cannot reveal the signal.

In equilibrium, the investigation effort of a media outlet depends on the expected return of investigation, which is determined by the outlet's belief about the type of the principal. This belief is based on the prior and on the internal capture strategy of the principal. Moreover, if the principal makes an offer, the transfer proposed matches the outside option of a media outlet: in the internal capture stage, the principal proposes the expected continuation payoff of the media outlet if it is not captured. In the external capture stage, if an outlet has

received a signal, the principal offers the publication payoff in exchange for its suppression. We uncover how the likelihood of the different forms of media capture depends on the parameters of the model, and in particular on the prior about the type of the principal.

First of all, if the prior probability that the principal is corrupt is low, media outlets have little interest in investigating. In such a case, the principal prefers to let the media outlets investigate freely and, potentially, capture them ex-post in the unlikely case that they receive a signal. This mechanism leads to the first key prediction of our paper: internal capture increases in the prior probability that the principal is corrupt.

Turning to external capture, the prior has two effects. Firstly, investigation efforts increase in the prior, which in turn increases the probability of receiving a signal that the principal has to suppress with external capture. However, a high prior also increases internal capture, which renders external capture unnecessary. The first effect dominates if and only if the media market is sufficiently large: the principal cannot internally capture enough media outlets to compensate for the increase in investigation efforts. This leads to the second key prediction of the model: the probability of external capture increases with perceived corruption if the media market is large enough and decreases otherwise.

In the second part of the paper, we study empirically the effect of perceived corruption on the choice of media capture strategy. The positive relationship between media capture and corruption is well established, but the causal relation is hard to identify. Hence, one of our contributions is to use the revelation of the Panama Papers as an exogenous shock to perceived corruption. The Panama Papers were released in 2016 by the International Consortium of Investigative Journalists[3]. It consisted of a leak of documents that explicitly mentioned individuals or firms involved in tax avoidance schemes. It triggered a worldwide scandal which received considerable coverage; nevertheless, this coverage varied across countries depending on the number of entities mentioned in the leaks. We exploit this variation in exposure to the shock.

Our identification strategy consists of two difference-in-differences estimations. Firstly, we establish that the Panama Papers did affect the perception of corruption and, more importantly, that the perception increased more in countries exposed to the Panama Papers after the revelations. Our estimates suggest that perceived corruption, measured by the Corruption Perception Index, increased by 1.2 percentage point – from an average score of 57 on a scale from 0 to 100 – in exposed countries after the shock.

We then discuss the suitability of the shock. We show that the revelation of the Panama

---

[3]The International Consortium of Investigative Journalists is a US-based nonprofit global network of reporters and media organizations created in 1997 with the goal of gathering outlets and journalists from all parts of the world to investigate cross-border stories. As of 2020, it includes 249 journalists from 90 countries.

Papers was indeed associated with an increase in interest in topics related to tax avoidance. Two mechanisms would fit our theoretical framework. The disclosure of new information on the magnitude of tax avoidance could trigger a rational updating of beliefs on corruption; alternatively, tax avoidance might receive more attention even if the existence of this practice was known before. However, an increase in actual corruption would undermine our identification assumption. We show that this channel is not at play in the direct aftermath of the revelations.

Then, we turn to the effect of perception of corruption on the forms of media capture. We use survey data provided by Reporters Without Borders, which cover 187 countries from 2012 to 2018. It is the most comprehensive cross-country database that distinguishes between the different forms of capture. We focus on the most salient examples of the two types: for internal capture, we consider conflicts of interest with media owners and the share of owners who have interests in other economic sectors. External capture is proxied by outside payments and in-kind benefits to journalists[4].

We use a second difference-in-differences and measure the effect of exposure to the Panama Papers on the prevalence of internal and external capture. We find a larger increase of internal capture in exposed countries, which confirms our first theoretical prediction. The share of owners with interests in other economic sectors increased by 9 percentage points in countries affected by the shock, compared to a baseline share of 68%.

To test our prediction on external capture, we use the number of television channels per capita as a proxy for the size of the media market. In line with our second theoretical prediction, the shock has a significantly larger effect on external capture in countries where the number of stations is above the median of the sample.

Finally, we conduct a series of robustness checks and explore the other channels through which the Panama Papers could have affected the media market. In addition to the change in perceptions of corruption that we document, the shock could have increased investors' interests in investigative journalism, which could result in the acquisition of media outlets and be misinterpreted as internal capture. Using data on advertising revenues and circulation of newspapers, we show that this mechanism is unlikely to drive our results. Moreover, the shock has similar effects on media capture in countries where no media outlet was involved in the International Consortium of Investigative Journalists, which suggests that the result is not driven by learning on investigative journalism capacity. We conclude that the effect of the shock on media capture is mainly driven by a change in perceived corruption and, thus,

---

[4]The timing of offers is the key element of our theory. Yet, some forms of capture – for instance through advertising – might occur at various stages of the investigation process. We look at other capture strategies, but the main analysis focuses on the forms that can be clearly associated with one of our two categories.

that our empirical findings support our theoretical predictions.

Literature review

If the substitution between the different strategies of media capture has received little attention, the various forms of capture are separately well documented (see Enikolopov and Petrova 2015 or Stiglitz 2017 for a review). For instance, Gilens and Hertzman (2000), Djankov et al. (2003), Dyck et al. (2008), Durante and Knight (2012), DellaVigna and Hermle (2017), and Fize (2020) show that media owners use their influence on the newsroom to affect the coverage of the news. Reuter and Zitzewitz (2006), Di Tella and Franceschelli (2011), Germano and Meier (2013), Szeidl and Szucs (2021), Beattie (2020) and Beattie et al. (2020) demonstrate that advertisers also manipulate the press and affect the tone of coverage. Moreover, direct payments to journalists affect media coverage, as documented in McMillan and Zoido (2004), Bignon and Miscio (2010) and Bignon and Flandreau (2011). Furthermore, Stanig (2015) and Gratton (2015) explain that the threat of a libel trial can deter publication. Finally, Eraslan and Özertürk (2018) document that politicians can influence coverage by selectively granting access to journalists.

The empirical part of our paper focuses on the relationship between corruption – or its perception (for the difference between the two, see Olken 2009) – and the freedom of the press. Brunetti and Weder (2003), followed by Chowdhury (2004) and Freille et al. (2007) document a negative correlation between press freedom and corruption. Those papers claim that the direction of causation goes from higher press freedom to lower corruption, but the correlation between the two variables could also be driven by reverse causality or by a simultaneous effect of unobservable variables, as explained in Prat and Strömberg (2013). Snyder and Strömberg (2010) establish a causal impact of press coverage on the performance of politicians. However, the effect of perceived corruption on media capture has received little attention and our paper is, to our knowledge, the first to investigate this question.

Besides, a contribution of our paper is to use the revelation of the Panama Papers as an exogenous shock to perceived corruption at the country level. Zucman (2013) estimates that 8% of the global financial wealth of households was held in tax havens in 2008. However, the span of tax evasion is still very opaque (Alstadsæter et al., 2019), and the Panama Papers disclosed new information on this practice. For instance, at the firm level, O'Donovan et al. (2019) document that the leaks revealed the existence of offshore vehicles to outside investors.

Our theoretical framework builds on the model of Besley and Prat (2006). Our paper aims at studying the interactions between the different possibilities of capture and we extend their main setup in two directions. Firstly, the probability that a media outlet receives a

signal depends on the effort of investigation[5]. Secondly, the principal chooses between two means of capture: suppressing the signals as in Besley and Prat (2006) or capturing media outlets before they investigate.

Our model assumes that a corrupt principal is hurt if a signal about his type is published. Alternatively, Gehlbach and Sonin (2014) propose a model of Bayesian persuasion where the utility of the principal depends on the posterior belief of citizens, who base their assessment on the report of the media and on the capture strategy of the principal. In a persuasion setup, the principal's intervention in the media market needs to be credible. This implies that the principal's ability to bias the news is higher when citizens trust him, as trustful citizens are more prone to be persuaded by a biased message[6]. In Appendix B.3, we introduce Bayesian persuasion in our model. If perceived corruption is large enough, our main prediction also reverses. Thus, the empirical analysis supports the model without Bayesian persuasion and we assume that the principal can only be punished if the media explicitly inform citizens that the principal is corrupt.

Also important in our model is the competition between media outlets. In line with Cagé (2019), media outlets investigate less when they have competitors because payoffs are split, which, in our paper, implies that competition favors delayed capture: the principal has an incentive to keep many media outlets free at the investigation stage because it reduces individual investigation levels[7].

Finally, a strand of the literature analyzes how a regime influences citizens' beliefs to prevent a revolution (Edmond 2013, Shadmehr and Bernhardt 2015 or Ananyev et al. 2019). Dziuda and Howell (2021) propose a partisan model of political scandals where parties strategically communicate on past misbehavior of a politician. Guriev et al. (2019) show that 3G coverage makes citizens more sensitive to corruption of governments and Guriev and Treisman (2019) consider a broader setup where the type of the regime can be revealed to citizens by an informed elite or a media outlet. Moreover, other papers like Corneo (2006) and Petrova (2008) focus on the interactions between citizens and media in a setup where citizens choose the level of redistribution. They show that the media can be captured by the rich to promote lower taxation rates.

The remainder of the paper proceeds as follows. We introduce our theory of media

---

[5]In Extension C, Besley and Prat (2006) also consider endogenous monitoring, but they focus on a unique capture strategy.

[6]Other papers studying media bias include Mullainathan and Shleifer (2005) and Gentzkow and Shapiro (2006). Gitmez and Molavi (2018) propose a bayesian persuasion model with heterogenous priors on the receiver side.

[7]Besley and Prat (2006) and Gehlbach and Sonin (2014) also consider the effect of n, but their modeling assumptions do not allow for strategic interactions between media outlets.

capture in Section 2. The empirical strategy is discussed in Section 3. Section 4 presents the empirical results. Section 5 concludes.

2   Model

2.1   Environment

We build on Besley and Prat (2006) and consider a model with a principal and a media market consisting in n media outlets indexed by i. The principal is either corrupt or honest: $\theta \in \{c, h\}$ with $P(\theta = c) = \gamma$.[8] In Appendix B.1, we discuss an extension where the principal privately chooses his type at the beginning of the game. Media outlets decide how much to investigate in order to learn the type of the principal: each media outlet chooses $q_i \in [0, 1]$ at cost $c(q_i)$, with $c(0) = 0$ and $c(.)$ increasing and convex. The detection of the principal's type depends on the realization of a random variable $v \sim U[0, 1]$ and media outlet i observes the type of the principal if $q_i \geq v$. v is commonly drawn for all media outlets and we assume that the principal observes which media outlets have received the signal. In Appendix B.2, we consider an alternative model where signals are independent.

Receiving and publishing a signal revealing that the principal is corrupt increases the revenue of the media outlet by a/m, where m is the number of media outlets publishing the signal. a represents the increase in sales or advertising revenues generated by the additional audience attracted by the signal. It also includes journalists' status and career prospects. To guarantee that media outlets choose an interior level of investigation, we suppose $c'(1) \geq a$.

A corrupt principal gets a negative payoff r if his type is revealed, which happens if at least one media outlet receives and publishes a signal. r captures the reaction of citizens[9]. In the main model, citizens have to observe a signal to punish the principal. In Appendix B.3, we draw on the Bayesian persuasion literature and propose an extension where the punishment depends on citizens' posterior belief. An honest principal is not affected by the revelation of his type and never engages in media capture.

As a result, we can focus on the corrupt type. Crucially, focusing on the corrupt principal allows us to derive predictions with respect to perceived corruption, which is the key dimension of the paper: it amounts to fixing the type of the principal ($\theta = c$) and analyzing how the outcome varies with media outlets' prior γ. In Appendix B.1, we account for

---

[8]In Besley and Prat (2006), the principal is an incumbent politician seeking reelection, but the model also applies to firms (as in Gilens and Hertzman 2000), businessmen (Dyck et al. 2008) or any actor who could manipulate the news.

[9]In an election, citizens elect the challenger if the incumbent is corrupt, as in Besley and Prat (2006) or in Guriev and Treisman (2019). In the case of a firm, consumers can buy from a different brand if they learn that the firm has been involved in some wrongdoings.

the probability that the principal is corrupt, which allows to compare actual and perceived corruption. We return to this point in the discussion at the end of this section.

We depart from Besley and Prat (2006) and consider two types of media capture. Firstly, the principal can use internal capture: at the beginning of the game, he proposes $t^I = \{t_1^I, ..., t_n^I\}$, where t_i^I is the offer to media outlet i. If media outlet i accepts the offer, the principal takes control of its investigation decision and sets $q_i = 0$. Offers and acceptance decisions are simultaneous and observed by all media outlets. Moreover, the principal can make a second offer $t^E = \{t_1^E, ..., t_n^E\}$ after the realization of the signals. We call this strategy external capture, and media outlet i suppresses the signal if it accepts the offer. In our main model, transfers must be positive. In Appendix B.4, we consider an extension where external capture can take the form of a negative payoff for the media outlets that publish, which represents threats and violence against journalists.

To sum up, the timing of the game is as follows:

1. *Internal capture.* The principal offers $t^I = \{t_1^I, ..., t_n^I\}$, media outlets simultaneously accept or reject their offer.
2. If media outlet i has not been internally captured, it chooses investigation q_i.
3. *External capture.* The principal learns the realization of the signals and offers $t^E = \{t_1^E, ..., t_n^E\}$, media outlets simultaneously accept or reject their offer.

2.2 Results

We solve the model backward and start with external capture. Depending on the number of media outlets that have received a signal, the principal either makes an offer that guarantees that all signals are suppressed, or lets the information out and pays the punishment cost r. When all media outlets are captured, the outside option of a media outlet which has received a signal consists in rejecting the offer and being one to publish. Thus, the offer of the principal must match the value of this outside option and he offers $t_i^E = a$ to all the media outlets which received the signal[10]. Hence, if at most $s = \lfloor \frac{r}{a} \rfloor$ signals have been received, the principal pays to suppress it. Otherwise, he lets the information be released and the media outlets which received a signal equally split the publication payoff a.

We now consider the investigation decisions. Media outlets can be ranked with respect to their investigation effort: $q_1 \geq ... \geq q_n$. As in Besley and Prat (2006), the equilibrium is described by the following lemma and is unique up to a renumbering of media outlets:

---

[10]If indifferent, we assume that media outlets accept the offer.

Lemma 1. *Let* $s = \lfloor \frac{r}{a} \rfloor$. *In equilibrium,* $q_1^* = ... = q_s^* = c'^{-1}(P[\theta = c|t^I]a)$ *and for all* $i \geq s+1$, $q_i^* = c'^{-1}(P[\theta = c|t^I]a/i)$.

If $v \in [q_1, q_{s+1}]$, the s media outlets with the highest investment level receive a from the principal: as v is common to all media outlets, exactly s signals are received and all of them are suppressed by the principal. The F.O.C. implies that these media outlets exert identical investigation efforts, equal to what a monopolist would choose. For $i \geq s+1$, media outlet i gets a/k if $v \in [q_k, q_{k+1}]$ for $k \geq i$. In such a case, the principal cannot suppress all signals. He lets media outlets publish and share the publication payoff. Thus, media outlets belong to two categories: s media outlets choose high investigation efforts and expect to get the full publication payoff with some probability under the form of an ex-post transfer from the principal. The other $n - s$ media outlets form the tail of the media market: they choose lower investment levels and share the publication profits when they receive a signal.

To simplify the notation, we assume in what follows that $n > s+1$ so that there are both media outlets with the high investigation level and in the tail. With regard to the principal, we can distinguish three cases. If $q_1^* < v$, no signal is received and the type remains secret. When $q_{s+1}^* < v \leq q_1^*$, exactly s media outlets receive the signal. The principal uses external capture to suppress it and pays $a \times s$. Lastly, if $v \leq q_{s+1}^*$, suppressing all signals is too expensive; the signal is revealed and the principal gets $-r$.

Now, we study internal capture and first intuitively describe the form of the equilibrium. Because of the signaling aspect of the game, the principal mixes between the early capture of a specific number of media outlets, and the waiting strategy consisting in $t_i^I = 0 \ \forall i$. If the corrupt principal always used internal capture, the absence of offers would reveal that the principal is honest ($\theta = h$) and imply no investigation of media outlets. Thus, the corrupt principal would deviate to imitate the honest type. Conversely, if the principal never used internal capture, media outlets would not update their prior. If the prior is large enough, media outlets would choose high investigation efforts, and the principal would prefer an early agreement: as we will see, early capture increases the principal's bargaining power because the investigation costs are not already paid. We now characterize the equilibrium offers formally.

The number of internally captured media outlets depends on the size of the media market. When n is small enough, the principal either waits ($t_i^I = 0 \ \forall i$) or internally captures all media outlets. An offer t_i^I is accepted if it is at least equal to the expected payoff of media outlet i if it remains free. Thus, when all media outlets are internally captured, each of them must receive the expected payoff of a monopolist.

Moreover, media outlets use t^I to update their belief about the type of the principal and adjust their investigation effort, which is a crucial difference with Besley and Prat (2006).

In particular, positive offers at the internal capture stage reveal that $\theta = c$, as a principal of type $\theta = h$ never makes offers. As a result, the principal must offer $t_i^I = \pi_1^{\theta=c} \ \forall i$ to capture all media outlets internally, where:

$$\pi_1^{\theta=c} = \max_{q_i}\{aq_i - c(q_i)\}.$$

This expression highlights a key aspect of internal capture: media outlets accept smaller offers at an early stage because they anticipate saving the cost of investigation. Instead, suppressing a signal after its realization is more expensive because the sunk cost of investigation is already paid. Thus, early agreements reveal the type of the principal, but also increase his bargaining power.

For larger values of n, the principal might buy the tail of least productive media outlets, i.e, internally capture media outlets $i \geq s + 1$. Each media outlet captured receives its deviation payoff, noted $\pi_{s+1}^{\theta=c}$, which corresponds to the payoff of media outlet $s + 1$ when $n = s + 1$ and $\theta = c$. In such a case, the principal relies on external capture to control the s media outlets that remain free.

Finally, for n large enough, the principal never engages in internal capture because it would imply buying too many media outlets. Formally:

Lemma 2. *There exists n_0 and n_1 such that[11]:*

- *if $n < n_0$, the principal mixes between $t_i^I = 0 \ \forall i$ and $t_i^I = \pi_1^{\theta=c} \ \forall i$,*
- *if $n_0 \leq n < n_1$ the principal mixes between $t_i^I = 0 \ \forall i$ and $\{t_i^I = 0 \ \text{for } i \leq s, t_i^I = \pi_{s+1}^{\theta=c} \ \text{for } i > s\}$,*
- *if $n \geq n_1$, the principal offers $t_i^I = 0 \ \forall i$.*

Now that we have identified the offers that the principal can make, we investigate the probability of internal capture and study how it depends on the prior about the type of the principal γ. We note the probability that the principal uses internal capture x. The equilibrium mixing probability x^* results from the comparison of the expected cost of $t_i^I = 0 \ \forall i$ with the cost of early capture.

First of all, notice that the cost of the internal capture strategy defined in Lemma 2 is independent of x and γ: following a positive offer, media outlets perfectly infer that $\theta = c$ and do not use the prior nor the mixing probability of the principal.

Instead, if $t_i^I = 0 \ \forall i$, media outlets use Bayesian updating to form their belief about θ:

$$P[\theta = c | t_i^I = 0 \ \forall i] = \frac{\gamma(1 - x)}{\gamma(1 - x) + 1 - \gamma}.$$

---

[11]$s + 1 < n_0 < n_1$ is not satisfied for all parameters.

Thus, investigation efforts, as well as the expected cost resulting from no internal capture offer, are increasing in γ and decreasing in x. A large prior increases the incentive to investigate, while a large probability of internal capture reduces the probability that $\theta = c$ if $t_i^I = 0 \; \forall i$. This implies:

Proposition 1. *The equilibrium probability of internal capture x^* is increasing in the prior probability that the principal is corrupt γ.*

When the principal is expected to be corrupt, media outlets choose high investigation efforts if they are not captured. Thus, the principal is more likely to capture media outlets ex-ante when γ is large. Early offers increase his bargaining power, as it spares the investigation cost. Moreover, revealing his type is less damaging when the prior is already high. Instead, when γ is low, uncaptured media outlets choose low investigation effort and are unlikely to receive a signal. Then, $t_i^I = 0 \; \forall i$ is more attractive for the principal[12].

Using a related reasoning, our model can address the effect of competition on the likelihood of internal capture. We have:

Corollary 1. *The equilibrium probability of internal capture x^* is decreasing in the number of media outlets n.*

This result is driven by two mechanisms. Firstly, the number of media outlets to be captured in order to affect the principal's continuation payoff is mechanically increasing in the size of the media market. Thus, an increase in n raises the cost of internal capture and makes it less attractive for the principal. Secondly, media outlets investigate on average less in a competitive market, which is in line with the evidence provided by Cagé (2019). Some of them choose low investigation levels because they expect to share the publication payoff if they receive the signal. Therefore, the principal wants to keep the market competitive for as long as possible[13].

Furthermore, we discuss the effect of the prior γ on the probability of external capture, which we note P^E. This probability depends on the strategy of internal capture adopted,

---

[12]In an alternative model with the simplifying (although arguably less realistic) assumption that the principal is not informed about his type, offers do not convey any information about the principal's type. In such a case, the costs of external and internal capture both increase in γ. Additional restrictions on the cost of investigation are needed to conclude on the effect of γ.

[13]For instance, in response to the capture of $i = 1$, all media outlets from the tail increase their investigation effort. If the media outlets received independent signals, removing an outlet would increase the effort of all others. We explore such a setup in Appendix B.2.

which is itself determined by n through Lemma 2:

$$
P^E = \begin{cases}
(1 - x^*)(q_1^* - q_{s+1}^*) & \text{if } n < n_0 \\
(1 - x^*)(q_1^* - q_{s+1}^*) + x^* q_1^{\theta=c} & \text{if } n_0 \leq n < n_1 \\
q_1^* - q_{s+1}^* & \text{if } n \geq n_1
\end{cases}
$$

When $n < n_0$, internal capture implies buying all media outlets. In such a case, external capture can only take place if media outlets have not been internally captured and if $v \in [q_{s+1}^*, q_1^*]$. Moreover, if $n_0 \leq n < n_1$, external capture can be observed even when the principal has used internal capture: there remains exactly s free media outlets, which choose investigation effort $q_1^{\theta=c} = \arg\max_{q_i}\{aq_i - c(q_i)\}$ and get externally captured if $v < q_1^{\theta=c}$. Finally, if $n \geq n_1$, internal capture is never used and media outlets are externally captured if $v \in [q_{s+1}^*, q_1^*]$.

The prior γ affects P^E through two channels. Firstly, it determines the probability that s signals are received if media outlets are not captured, which is the condition to observe external capture and corresponds to the size of the interval $[q_{s+1}^*, q_1^*]$. Secondly, γ has a positive effect on the probability of internal capture x^*, which affects P^E differently as n varies. In particular, if $n < n_0$, the increase in internal capture results in less external capture because media outlets have agreed not to investigate, and thus cannot receive a signal.

To keep the model tractable, we impose the following restriction, which guarantees the regularity of the growing rate of investigation efforts:

Assumption A: $c'(q) = \kappa q$, $\kappa \in \mathbb{R}^+$.

We obtain:

Proposition 2. *Under Assumption A, there exists a number of media outlets $\bar{n}$ such that the probability of external capture $P^E(\gamma)$ is decreasing in the prior γ if $n < \bar{n}$ and increasing otherwise.*

To illustrate the main predictions, we simulate the model and plot the probabilities of internal and external capture for $n < \bar{n}$ and $n > \bar{n}$ in Figure 1.2. When the media market is small, the larger control of journalists' agendas following an increase in γ dominates the potential increase in investigation of free journalists. Thus, the probability of external capture decreases because investigation has been deterred at an early stage. However, when the market is large enough, the principal cannot sufficiently increase internal control because it implies buying too many media outlets. The rise in investigation efforts dominates: media outlets are more likely to receive s signals, which are suppressed by the principal through external capture.
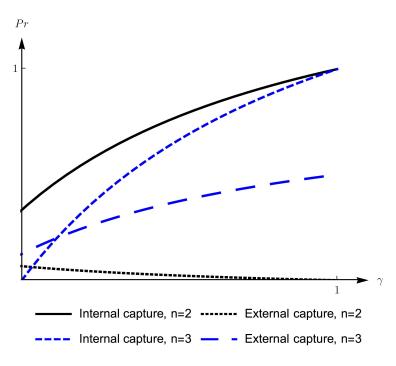
Figure 1.2: Probability of capture

Notes: Probability of internal and external capture as a function of γ for $n = 2$ and $n = 3$. $a = 1/2$; $r = 3/2$ and $c(q) = q^2/2$. Plotted for $\gamma \in [1/2, 1]$ to have a positive probability of internal capture.

2.3   Discussion

The key predictions of our model relate to the effect of perceived corruption on media capture and can be summarized as follows. Firstly, we expect internal capture to increase with perceived corruption. An early agreement is preferred by the principal if media outlets would exert high investigation efforts, which is the case if they initially believe that the principal is corrupt. Turning to external capture, we expect a positive effect of perceived corruption if the media market is competitive, but a negative effect if it is less developed. This prediction is driven by two opposite forces: an increase in perceived corruption results in more investigation from media outlets, which in turn increases the necessity of external capture. On the other hand, perceived corruption increases internal capture, which renders external capture unnecessary. When the media market is small, the second effect dominates, which explains our result on external capture.

In Appendix B, we discuss the robustness of our results to alternative modeling assumptions and present some extensions. First of all, the type of the principal is endogenously chosen in Appendix B.1. The principal chooses whether to engage in corruption at the beginning of the game, which provides him with a positive payoff but induces a risk of being punished. This extension aims to explicitly distinguish actual corruption from perceived

corruption. The former is the true probability that the principal is corrupt while the latter is the belief of the media market, which corresponded to γ in the main model. We find that our key results on perceived corruption extend to actual corruption. In particular, the probability of internal capture increases with the return of corruption. Moreover, under Assumption A, the probability of external capture increases with the payoff of corruption if the media market is large enough. However, when the market is small, media outlets are always internally captured and there is no external capture.

Furthermore, we show that our predictions do not depend on the assumption that signals are correlated. Appendix B.2 presents a model where media outlets receive independent signals: each media outlet $i \in \{1, 2\}$ receives a signal with probability q_i. As in the main model, the probability of internal capture is increasing in perceived corruption and the probability of external capture is more likely to decrease if $n = 1$ than if $n = 2$.

In our main model, media outlets use the offers to update their beliefs about the principal's type, but citizens do not. We made this assumption for two reasons: firstly, citizens arguably pay less attention to ownership changes than journalists[14]. Secondly, citizens are unlikely to punish the principal only because he bought a media. This could be modeled as in Ananyev et al. (2019), where one of several protest strategies is efficient. This strategy is initially unknown and can be revealed by the media. Without this information, citizens do not protest. In the terms of our model, citizens need the signal to learn the actual wrongdoing of the principal and punish him.

Alternatively, Appendix B.3 explores a Bayesian persuasion setup where the principal is punished if citizens believe that he is corrupt with a large enough probability. Citizens' posterior belief depends on the report of a (unique) media outlet and on the principal's capture strategy. In such a case, the principal's strategy is subject to a credibility constraint. If the prior is small, this constraint does not bind and the two models are equivalent. However, if the prior is large enough, excessive media capture could backfire: the absence of a signal is not sufficient to persuade citizens not to punish the principal and he needs to reduce media capture to remain credible. Thus, if the prior is large, a Bayesian persuasion approach reverses our main prediction: an increase in perceived corruption tightens the credibility constraint and decreases the probability of internal capture[15]. Moreover, this setup implies that the probability of external capture is always decreasing in γ. Both predictions of the

---

[14]Documenting the acquisition of TV stations in the U.S. by Sinclair, Mastrorocco and Ornaghi (2020) claim that the change is not salient enough to be actively noticed by viewers, at least in the short run: "after acquisitions, stations maintain their call sign, network affiliation, and news anchors." Thus, viewers do not account for the change in bias and are indeed manipulated by the new owners (Mastrorocco and Ornaghi 2020, Miho 2020).

[15]Notice that a similar mechanism exists in other models of media bias, like Gehlbach and Sonin (2014): the media bias is increasing in the probability that the state is favorable to the principal.

Bayesian persuasion model are at odds with the empirical evidence of Section 4.

Furthermore, Appendix B.4 studies the consequences of violence and intimidation of journalists, which we model with negative external capture offers. The probability of a negative offer depends on the protection of journalists and affects the investigation decisions of media outlets. The possibility of a violent suppression of the signals triggers a decrease in investigation, known as the "chilling effect" (Stanig 2015, Gratton 2015). Moreover, if the protection of journalists is low, there is an equilibrium where investigation efforts are strategic complements: the principal cannot use violence if a large enough number of media outlets receive the signal, implying that investigation is more profitable if other media outlets investigate as well. In such a case, a coalition of media outlets choose a common level of investigation in order to receive and publish the signal simultaneously without being threatened by the principal. In response, the principal internally captures a limited fraction of the media market, so that the uncaptured media outlets are not enough to form a coalition which would avoid a violent suppression of the signal.

Finally, we would like to discuss further two assumptions of the model made for tractability. Firstly, the only purpose of media capture is to silence the media, which implies that internal capture perfectly reveals the principal's type to media outlets. A similar mechanism is in Guriev and Treisman (2019), where only a corrupt principal engages in repression. Yet, not every media acquisition is nefarious. An attempt to incorporate this consideration is adding a third type of principal (media tycoon) who makes offers without hiding a scandal, for instance because he sees media outlets as valuable investments. This alternative modeling assumption makes the signaling part more involved. In particular, the updating of media outlets' beliefs following an internal capture attempt depends on which offers would be made by the tycoon. Secondly, we follow Besley and Prat (2006) and assume that the principal is punished if at least one signal is published. This simplifies the market for news to the extreme and leads to the stark prediction that the principal captures the whole media market in the first case of Lemma 2. Alternatively, if audiences are segregated, the disclosure of the signal by one media outlet would not affect all citizens. Then, the principal could only capture specific media outlets in order to prevent the disclosure of the signal to some sections of the population.

3   Empirical strategy

In the remainder of the paper, we study empirically the relationship between the perception of corruption and strategies of media capture. Estimating the relationship between those dimensions raises several endogeneity concerns. While previous studies have docu-

mented a negative correlation between corruption and the freedom of the press (Brunetti and Weder 2003, Besley and Prat 2006, Freille et al. 2007), the direction of the causation is hard to establish.

We propose using the disclosure of the Panama Papers as a shock to perceived corruption. The revelation of the Panama Papers started with the leak of documents from Mossack Fonseca, a Panamanian law firm and corporate service provider. The documents, which were sent anonymously to the German newspaper *Süddeutsche Zeitung* in 2015, reported cases of offshore entities hosted in tax havens, held by individuals from throughout the world. This newspaper shared the documents with the International Consortium of Investigative Journalists, made up of journalists from 70 countries. This group privately researched on the offshore entities mentioned in the leaked documents and disclosed a massive and global tax evasion scandal in April 2016.

In the framing of our model, we believe that the shock corresponds to a change in the prior γ: it does not affect the actual type of the principal (at least in the short run), as his decision to engage in tax evasion was already made. However, we expect the revelations to trigger an increase in media outlets' prior, in particular in countries where the scandal was large.

After we present the data that we use, we establish that the perception of corruption increases more in countries exposed to the Panama Papers after the revelations and discuss the validity of the shock. We look at the effect on media capture and test our theoretical predictions in Section 4.

3.1 Data

Perceived corruption

Our main measure of perceived corruption is the Corruption Perception Index published by Transparency International. This index is computed every year for more than 180 countries and ranges from 0 (no corruption) to 100 (high corruption). All summary statistics are in Table 1.1.

Even if the Corruption Perception Index is primarily designed to measure the perception of corruption in the public sector, it aggregates multiple sources and reflects the perception of corruption broadly defined, including the prevalence of tax avoidance (Transparency International, 2018). In particular, it should be a satisfactory proxy for the beliefs of the actors in the media sector.

Table 1.1: Summary statistics

	Mean	SD	Min	Max	Count
Media landscape					
Internal capture					
Owners with interests in other economic sectors	5.29	2.24	0	10	953
Conflicts of interests with owners	6.84	2.19	1	10	953
External capture					
Advantages for coverage	7.50	1.52	2	10	941
Journalists receive money from other sources	6.78	1.95	2	10	953
Press Freedom Country Score	3.43	1.70	1	9	1,000
Perceived corruption					
Corruption Perception Index	57.15	19.68	8	92	1,203
Panama Papers					
# Offshore entities	875.49	4,224.51	0	37,911	177
# Offshore entities if # > 0	1,192.02	4,895.76	1	37,911	130

Notes: Media landscape - Source: Reporters Without Borders Experts Survey. The sample contains one observation per country for each year: 169 countries between 2012 and 2018 except 2014. When several experts are surveyed for a given $year \times country$, we average their answer to get a balanced panel. *Global country score*: normalized to range from 1 (completely free) to 10 (not free at all). Internal and external capture indexes lie between 1 (best conditions, no pressures) and 10 (worst conditions). *Conflicts of interest with media owners*: Are media owners' conflicts of interest frequently the cause of journalists' self-censorship? *Owners with interest in other sectors:* What proportion of general-interest media is owned by companies with other interests in non-media sectors of the economy? (One unit corresponds to 10%). *Advantages for coverage:* Do some journalists receive invitations to luxury events, press trips and other benefits of all kinds that would weaken their publications' objectivity? *Journalists receive money from other sources:* Are journalists sometimes paid by someone other than their regular employer in order to influence what they write? Perceived corruption - Source: Transparency International (2013- 2018). The initial scale is reversed for interpretation and ranges from 0 (no perceived corruption) to 100 (most widespread corruption). Panama Papers - Number of offshore entities excluding tax havens. *Offshore entities if # > 0:* includes countries mentioned at least once in the Panama Papers.

Panama Papers

We use the list of the offshore entities mentioned in the leaked documents to measure country exposure to the Panama Papers.

Each entity is associated with its country of origin and we measure exposure with the

number of entities mentioned by country[16]. The average is 875 and the distribution, plotted in Figure C.1 in Appendix D, is very right-skewed.

The maps shown in Figures C.2 and C.3 illustrate the worldwide repartition of entities mentioned. About one third of countries were not exposed at all. Noticeably, developed countries tend to be less corrupt but more exposed to tax evasion behaviors, as revealed by the negative correlation plotted in Figure C.4.

Media capture

To measure media capture, we use a dataset provided by Reporters Without Borders. To our knowledge, this is the first data source that distinguishes the different forms of media capture at a cross-country level. Expert surveys are conducted in 180 countries every year since 2012, to the exception of 2014. Selected experts have to answer a wide range of questions related to the freedom of the press in their country.

We focus on the questions that best represent the two types of capture. For internal capture, we consider the share of media owners with interests in other economic sectors. If new owners enter the media market to influence journalists' agenda, we would observe an increase in this indicator. We also use the likelihood of conflicts of interest with owners, which reflects how much the agenda of the newsroom is influenced. This second variable should also respond positively if new owners enter the media market to manipulate the news. Yet, it can also be affected by a change in strategy of owners who were already in the market. Therefore, the effect we want to test might be best tracked by the first question. The exact wording of the questions can be found in Table 1.1.

Turning to external capture, we focus on in-kind benefits and external payments for coverage offered to journalists. We expect such transfers to take place after the investigation stage and to be given to the journalists who are about to publish on sensitive topics. Two alternative interpretations have to be discussed. Firstly, transfers to editors or to key journalists who decide which topics are investigated could be equivalent to internal capture. We cannot rule out that some experts interpreted the questions in such a way, but the statement explicitly focuses on payments to journalists. Moreover, if some very specific issues can only be investigated by specialized journalists, most journalists can investigate a large range of topics and, in a media outlet, a given topic can be investigated by several journalists[17]. Hence, bribing them individually would require specific and up-to-date knowledge

---

[16]Ideally, this measure should be completed with the notoriety of entities mentioned, which differed across countries and is likely to have affected exposure. Yet, such data are not available. The Panama Papers data can be accessed on the website of the International Consortium of Investigative Journalists: `https://offshoreleaks.icij.org/`

[17]For instance, on the selection of investigation topics at the New York Times, see Van Syckle, Katie, *How*

about the newsrooms and to hand in multiple bribes, making this option less efficient than a centralized internal capture strategy. Secondly, if the principal repeatedly interacts with journalists, external payments could also prevent future investigation[18]. However, even in a repeated setup, multiple individual payments would be dominated by a centralized control of the media. Thus, we expect in-kind benefits and external payments to target journalists about to release sensitive information, which in our model corresponds to external capture.

Finally, we consider the general score of the freedom of the press computed by Reporters Without Borders. This score aggregates the different dimensions encompassed in the survey.

In the remainder of the analysis, we focus on countries that are both included in the Transparency International and in the Reporters Without Borders datasets for more than 2 consecutive years. The final sample includes 169 countries.

3.2   The Panama Papers and perceived corruption

To test for the effect of the Panama Papers on the perception of corruption, we estimate the following difference-in-differences model:

$$y_{ct} = \lambda + \phi_t + \phi_c + \beta.After_t \times Exposure_c + \delta.After_t \times X_{ct} + \epsilon_{ct}. \tag{1.1}$$

The dependent variable y_{ct} is the Corruption Perception Index in country c in year t. $After_t$ is an indicator variable equal to one after the shock, i.e when $t \in \{2016, 2017, 2018\}$. $Exposure_c$ is a function of the number of entities from country c mentioned in the leaks. All specifications include year and country fixed effects ϕ_t and ϕ_c. To overcome possible correlations between country characteristics and post-2016 trends in perceived corruption, we interact $After_t$ with a set of country controls X_{ct}. Standard errors are clustered at the country level. Thus, the parameter of interest β is a difference-in-differences estimate measuring the effect of exposure to the Panama Papers on perceived corruption.

We estimate Regression 1.1 in Table 1.2. First, we focus on the extensive margin effect: $Exposure_c$ is an indicator variable equal to 1 if at least one offshore entity from country c is mentioned in the leaks. Without controls, the difference-in-differences estimate displayed in Column 1 shows that being mentioned in the Panama Papers triggers a 1.2 point increase in the Corruption Perception Index after the revelations. This coefficient is significant at the 5% level. In Column 3, we introduce a set of country characteristics – including logarithm

---

The Times Decides What to Investigate. The New York Times, March 20, 2019.

[18]In an infinitely repeated game, we can construct an equilibrium where journalists do not investigate because the principal threatens to exclude them from the payment of future bribes if they investigate, while journalists threaten to investigate if they do not receive the bribe. In such an equilibrium, it could be argued that bribes deter future investigation.

Table 1.2: The Panama Papers and perceptions of corruption

| | Corruption Perc. Index | | | |
	(1)	(2)	(3)	(4)
After × Exposure	1.214**		1.382*	
	(0.583)		(0.795)	
After × Exposure (log)		0.096		0.137
		(0.086)		(0.132)
Observations	1,133	1,133	1,045	1,045
Mean DepVar	57.90	57.90	56.56	56.56
Sd DepVar	19.85	19.85	19.43	19.43
Adj-R2	0.99	0.99	0.99	0.98
Year FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
After × Controls			✓	✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimation of Regression 1.1. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. The dependent variable is Transparency International's Corruption Perception Index. *Controls* include log GDP, growth rate over the period, population, and number of multinational companies based in the country.

of GDP, population, yearly growth and number of multinationals – that we interact with the post-treatment variable. The coefficient remains significant at the 10% level and the point estimate is even slightly larger, indicating that the different post-treatment trends are not driven by heterogeneous country characteristics between the treatment and the control group. We study the intensive margin in even columns and define $Exposure_c$ as the logarithm of the number of entities mentioned[19]. We find a positive but not significant relationship in Columns 2 and 4.

With regard to the magnitude of the effect, the estimates are around 1.2-1.4, which is

[19]To deal with zero values, we add one to the number of Panama Papers before the logarithm transformation.

not trivial given that they correspond to within-country variation. While ranging from 0 to 100, the Corruption Perception Index is very inertial[20] at the country level, as illustrated by the large R^2.

To compare the trends in perceived corruption between the treatment and the control group before the shock, we interact $Exposure_c$ with all indicator variables instead of using the post-treatment variable, and plot the estimates in Figure 1.3. The coefficients of the interaction terms are small and not significant for $t \leq 2015$, which confirms the absence of pre-trends before the shock and so the validity of the difference-in-differences approach. Moreover, we find positive and significant coefficients of similar magnitude for $t \in \{2016, 2017\}$. The effect is smaller and less significant in 2018, which suggests a fading of the effect.

We perform a series of robustness checks and report the results in Appendix D. First of all, Table D.1 shows that our results are robust to the exclusion of tax havens, namely Luxembourg, Switzerland and Hong Kong, which are clear outliers in the number of entities mentioned[21]. Moreover, to account for the fading of the effect, we exclude 2018 in Table D.2. This specification reinforces our results: for instance, at the extensive margin, we find a 1.6 point increase in the Corruption Perception Index after the shock. Finally, we replicate the specifications of Table 1.2 using the World Governance Indicators, an alternative measure of perceived corruption published by the World Bank[22]. The results shown in Table D.3 are consistent in terms of magnitude, but less significant.

3.3   Validity of the shock

We have established that the perception of corruption increased in countries exposed to the Panama Papers after the revelations. Now, we discuss the suitability of the shock to test our predictions on media capture.

Actors involved in media capture, in particular journalists and outsiders who want to manipulate the news, are very likely to be aware of the story. We use data from Google Trends to confirm the salience of this topic. In Figure C.5, we plot the evolution of the number of searches containing the words "Panama Papers" and several terms related to tax avoidance such as "shell companies" or "tax evasion". The first graph, based on the keyword "Panama Papers" exhibits a sudden jump in the number of queries after the revelation, which

---

[20]The average change in the score at the country level per year is of -.06 percent, with a standard deviation of 0.97.

[21]Other countries considered as tax havens – for instance those appearing on the European Union tax haven blacklist – are not included in the initial sample.

[22]The structure of the data as well as the objective of this indicator are similar to the Corruption Perception Index (Kaufmann et al., 2011). Besides the positive correlation between the indexes, the two sources differ in terms of methodology and measures of perceptions.
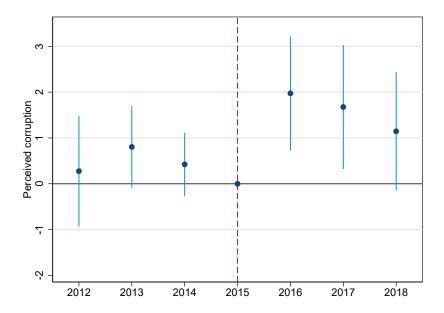
Figure 1.3: Pre-trends analysis

Notes: Estimated equation: $Corruption_{ct} = \lambda + \phi_t + \phi_c + \sum_{i=t, t \in \{2012, \dots, 2018\}} \beta_i . \phi_i \times Exposure_c + \epsilon_{ct}$, we report the coefficients and confidence intervals of the β_i. The dependent variable is Transparency International's Corruption Perception Index, *Exposure* is equal to 1 when country c is mentioned at least once in the Panama Papers. Standard errors are clustered at the country level. We omit the interaction with $t = 2015$ for identification. Confidence intervals are computed at the 95% level.

confirms that the revelation is indeed not anticipated. The surge in queries related to tax evasion visible in the second and third graphs also emphasizes that the shock triggers a rise in interest in these topics.

Several mechanisms can drive the impact of the shock on the perception of corruption. First of all, the Panama Papers could reveal hard information about tax avoidance. If experts surveyed for the Corruption Perception Index were not perfectly aware of the prevalence of this phenomenon in their country, they could rationally use this information to update their beliefs. Indeed, there remains a significant variation in the number of entities mentioned when we compare countries similar in GDP or population, as we can see in the maps showing the worldwide distribution of entities (Figure C.3), which suggests that the information contained in the revelation could not be perfectly predicted by country characteristics.

A second explanation relates to the biased perception of corruption, documented for instance in Olken (2009). Rizzica and Tonello (2015) show that the perception of corruption is affected by media coverage of corruption topics, even if it does not contain new information. The Panama Papers are likely to bring tax avoidance into the limelight, and the change in experts' opinions can result from an increased sensitivity to tax avoidance issues. This effect

would be stronger in exposed countries, as the Panama Papers receive more attention.

These first two explanations do not exclude each other. As we are interested in perceptions, the shock could be purely informational, behavioral or a combination of both. As long as journalists revise their perception of corruption similarly as surveyed experts do, this does not challenge the validity of the identification strategy.

However, a third possible explanation would undermine the interpretation of the results: the revelation of the Panama Papers could affect the actual level of corruption, for instance through a change in behavior of entities mentioned or a reaction of governments. O'Donovan et al. (2019) document that firms that were using their offshore entities illegally had to cease these activities after the publication of the leaks. At the firm level, they thus claim that the Panama Papers reduced corruption. In our theoretical framework, this would translate into a change in the number of corrupt principals instead of a change in capture strategy, a mechanism we theoretically describe in Appendix B.1.[23]

To test whether the impact on actual corruption can be generalized at the country level, we follow Guriev et al. (2019), who use the Global Incidents of Corruption Index of the International Monetary Fund constructed by Furceri et al. (2019). These data include a variable for actual corruption, which is the share of corruption-related terms in the country annual reports published by the Economist Intelligence Unit. On average, 0.27% of the reports relate to corruption[24]. In Table D.4, we estimate Regression 1.1 with objective corruption as the dependent variable. All coefficients are not significant and close to 0, both at the intensive and extensive margins, which suggests that the Panama Papers did not affect the actual level of corruption at the country level, at least in the short run.

4   Main empirical results

In this section, we look at the effect of exposure to the shock on the evolution of the different types of media capture. We first test our theoretical predictions. Then, we propose several robustness checks. Finally, we investigate some mechanisms that could also explain our empirical results.

---

[23]Notice that the results of O'Donovan et al. (2019) combined with Appendix B.1 cannot explain the results of Section 3.3: the Panama Papers trigger a decrease in the return of corruption. Thus, internal capture should decrease because the number of corrupt principals as well as their likelihood to use internal capture decrease, which is at odds with our empirical findings.

[24]The minimum and maximum values are 0 and 1.8%, the standard deviation is 0.3.

4.1 The Panama Papers and media capture

To test the effect of the Panama Papers on the different forms of media capture, our preferred approach is a second difference-in-differences model similar to Regression 1.1, using the measures of media capture as the dependent variables. In the robustness checks, we also implement an instrumental variable approach[25]. We report the results of the difference-in-differences regression in Table 1.3. For each outcome, we show the intensive as well as the extensive margin effects.

In Columns 1 and 2, we look at the share of media owners who have interests in other economic sectors. At the country level, being mentioned in the leaks leads to a 9% increase after the shock. This finding does not contradict our theory: after an increase in perceived corruption, internal capture becomes more attractive for the principal. The effect on the intensive margin is also significant: a 1% increase in exposure increases the share of external owners by 1 percentage point. In Columns 3 and 4, we consider conflicts of interest with media owners. The point estimate is also positive besides not significant.

In Columns 5-8, we turn to external capture. We find that the likelihood of in-kind benefits as well as external payments to journalists do not increase in our difference-in-differences specification. All estimates are not significant and close to zero. This result is consistent with the theory: Proposition 2 predicts that external capture should increase in countries where the media market is sufficiently large and decrease otherwise. These effects could cancel each other out, which would explain the coefficient.

The positive result on internal capture could be driven by a general shift in experts' perception. Given the salience of the Panama Papers, the shock could increase perception or awareness of corruption issues in affected countries. This explanation would imply a negative effect on the perception of the freedom of the press in general and, thus, an increase in all the dimensions of our surveys. This is not supported by the absence of significant results on external capture. We look at the aggregate index of the freedom of the press, which summarizes all dimensions included in the survey in Columns 9 and 10. We find that the effect on this outcome is not significant and very close to 0. This suggests that the result on internal capture is not spurious: experts do not become more pessimistic in general[26].

Our prediction on external capture is harder to test. It requires a measure of the number of news media outlets, which is not included in our primary data sources. We consider the number of television channels from the CIA's World Factbook[27]. We focus on the television

---

[25]Because of the absence of the year 2014 in the media capture data, this approach gives less significant results.

[26]The variables *Owners* and *Conflicts* pertain to the block "Environment and self-censorship" which consists in 18 variables, representing less than a third of the total score.

[27]https://www.cia.gov/the-world-factbook/. Number of television channels for 100,000 inhabitants

Table 1.3: The Panama Papers and media capture

| | Internal capture | | | | External capture | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
	Owners	Owners	Conflicts	Conflicts	In-kind	In-kind	Bribes	Bribes	Score	Score
After × Exposure	0.920***		0.244		-0.137		0.014		-0.369	
	(0.348)		(0.370)		(0.137)		(0.144)		(0.664)	
After × Exposure (log)		0.143***		0.031		0.015		0.034		0.058
		(0.051)		(0.044)		(0.023)		(0.021)		(0.089)
Observations	950	950	950	950	938	938	950	950	982	982
Mean DepVar	5.29	5.29	6.84	6.84	4.50	4.50	4.07	4.07	34.48	34.48
Sd DepVar	2.25	2.25	2.19	2.19	0.91	0.91	1.17	1.17	17.03	17.03
Adj. R2	0.53	0.53	0.60	0.60	0.49	0.49	0.65	0.65	0.98	0.98
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimation of Regression 1.1. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. Dependent variables: *Owners*: proportion of general-interest media owned by companies with interests in other economic sectors. *Conflicts*: likelihood of journalist self-censorship due to conflicts of interest with the owner. *In-kind*: non-monetary transfers to journalists to influence the coverage. *Bribes*: payments to journalists to influence the coverage. *Score*: Reporters Without Borders's aggregate score for the freedom of the press.

market because it is generally considered as more representative and unified at the country level than other media supports (Angelucci et al., 2020). Yet, a limitation of the number of media outlets – in particular television channels – is to include entertainment outlets with little informational content and no investigative capacity. These media outlets should play little role in the capture strategy of the principal. Moreover, we do not know how many channels are actually free to investigate. For instance, channels controlled by politicians in office are unlikely to reveal scandals implying them. Finally, we expect external capture to decreases if the number of outlets is sufficiently small, but the threshold above which the relationship reverses is not predicted by the model. Unfortunately, available data do not allow us to overcome these limitations. In the robustness checks, we propose an alternative approach and use a different proxy for the number of media outlets.

We depart from Regression 1.1 and interact our treatment variable – i.e. country exposure to the Panama Papers – with a categorical variable indicating whether the number of television channels per 100,000 inhabitants is above the median of the sample. In Table 1.4, we find that the revelations have significantly different effects on external capture in the two groups. In Column 1, the prevalence of in-kind benefits seems little affected at the extensive margin by the revelations if the media market is competitive, while it tends to decrease in less competitive markets. The effect is similar for bribes but not significant. Looking at the intensive margin, the point estimates suggest an increase in external capture for countries with more television channels. These results are consistent with our second theoretical prediction: in countries where the media market is sufficiently large, an increase in perceived corruption stimulates journalistic investigation, which leads to more ex-post pressures. In the other group of countries, this effect is compensated by the increase in internal capture.

Finally, we also estimate Regression 1.1 on the other types of capture mentioned in the introduction and report the results in Table D.5. The revelation of the Panama Papers has no effect on jailing of journalists and leads to more violence and retaliation, which is theoretically discussed in Appendix B.4. The Reporters Without Borders data do not investigate advertising pressures by private companies. With regard to state advertising, we find that it positively covaries with perceived corruption.

4.2   Robustness checks

We now perform a series of robustness checks to validate our results on media capture. First of all, we introduce the interaction between $After_t$ and the set of country controls (logarithm of GDP, population, yearly growth and number of multinationals) in Regression

---

are available for 107 countries. Mean is 0.64, standard deviation is 1.6 with a minimum of 0, a maximum of 8 and a median of .23.

Table 1.4: Heterogeneity in External capture

	(1)	(2)	(3)	(4)
	In-kind	Bribes	In-kind	Bribes
After × Exposure				
High Competition	-0.074	0.054		
	(0.229)	(0.298)		
Low Competition	-0.386*	-0.205		
	(0.212)	(0.305)		
After × Exposure (log)				
High Competition			0.092***	0.088***
			(0.030)	(0.032)
Low Competition			-0.011	-0.002
			(0.026)	(0.030)
Observations	596	606	596	606
Mean DepVar	4.51	3.95	4.51	3.95
Sd DepVar	0.91	1.15	0.91	1.15
Adj. R2	0.54	0.67	0.55	0.68
Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
pvalue: High=Low	.054	.097	0.00	.003

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimation of Regression 1.1, interactions with level of competition added. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. *High competition (resp. Low Competition)*: number of TV channels per 100,000 inhabitants above (resp. below) the median in the sample (m=.23).

1.1 with the measures of media capture as dependent variables. The results, reported in Table D.6, are similar to what we obtained in Table 1.3. This indicates that the effect on internal capture is not driven by the correlation between post-trend differences in country characteristics and the extent of capture. The results are also very similar in Table D.7, where we exclude 2018 from the analysis to account for the possible fading effect documented in

Figure 1.3. Turning to the results on external capture, they are also robust to the inclusion of post-shock controls, as shown in Table D.8.

Furthermore, we test whether our results could be driven by specific countries in Table D.9. In Panel A, we exclude tax havens from our sample (Luxembourg, Switzerland and Hong Kong), which are outliers in the number of entities mentioned in the leaks. An additional concern could be the influence of some countries on the publication process, even if the independence of the International Consortium of Investigative Journalists renders such manipulations less likely. In Panel B, we exclude the USA, China and Russia from the analysis[28]. With both sample restrictions, the results are very similar to Table 1.3.

We also estimate Regression 1.1 with an ordered logit and report the results in Table D.10. As in Table 1.3, we find a positive effect of the shock on internal capture while the general effect on external capture is smaller and not significant.

We then examine the parallel trends assumption in Table D.11, where we interact the treatment variable with all year indicator variables[29]. This specification mitigates the concern of possible pretrends: for the share of media owners who have interests in other economic sectors, none of the interaction coefficients is significant before the shock, both at the intensive and extensive margins, while almost all interactions are positive and significant for $t \in \{2016, 2017, 2018\}$. Turning to the other dimensions, few interaction coefficients are significant.

Moreover, we implement an alternative identification strategy and instrument the change in perceived corruption by the exposure to the Panama Papers. We report the estimation of a 2SLS model in Table D.12. In the first stage, we regress the Corruption Perception Index on the exposure to the Panama Papers and we use the predicted values to explain media capture in the second stage. Consistently with the results obtained with the difference-in-differences, the point estimates are positive for internal capture, close in magnitude to the previous estimates, and near 0 for external capture. However, the coefficients of the second stage are not significant. This lack of significance results could stem from data limitations of the Reporters Without Borders' dataset: the year 2014 is not available and has to be excluded for both stages. Then, it affects the significance of the first stage coefficient, i.e. the estimate of the Panama Papers on the Perceived Corruption Index[30]. Moreover, while the extensive margin treatment gives a satisfactory instrument (F=10.4), our second treatment (logarithm of the number of offshore entities) is too weak to be a satisfactory instrument and

---

[28]Results do not change if we exclude countries individually.

[29]The absence of data on media capture in 2014 prevents us from representing the pretrends as in Figure 1.3.

[30]While the estimates displayed in Table 1.3 are significant at the 5% level, the standard errors of the first stage estimates shown in Table D.12 are larger.

we cannot explore the robustness at the intensive margin with an instrumental approach.

Finally, we use an alternative proxy for the size of the media market to test the effect on external capture. Reporters Without Borders propose a breakdown of the sample in three groups based on the score of the freedom of the press[31]. We expect this partition to reflect, at least partly, the level of competition on the media market. Figure C.6 shows a positive correlation with the measures of competition constructed by Noam (2016), which are only available for 29 countries in 2011 and so can only be used to check the suitability of this approach[32]. We estimate Regression 1.1 in Table D.13 and interact the treatment variables with the freedom of the press categories. The results are consistent with Table 1.4: the effect on external capture is significantly larger in countries with a competitive media market. This alternative approach alleviates some concerns pointed out for the number of television channels, but raises possible endogeneity issues.

4.3   Alternative mechanisms

Finally, we turn to the possible alternative mechanisms that could also explain our results on media capture. The key assumption of the identification strategy is that the parameters that affect the freedom of the press – excluding the perception of corruption – do not evolve differently after and because of the revelation of the Panama Papers. Under this assumption, the country fixed effects, as well as the introduction of post-shock controls, adequately control for those factors in our difference-in-differences specification. We test two potential confounding channels: first, the Panama Papers could have revealed information on journalists' investigative capacity and, second, it could have affected the profitability of the media sector.

Investigative capacity

Before the revelation of the Panama Papers, a set of media outlets joined the International Consortium of Investigative Journalists to analyze the leaked documents. This investigation process might have affected the beliefs about journalists' ability and pushed investors to enter the media market. In line with what we found in Table 1.3, this effect could be stronger in exposed countries because the Panama Papers received more attention. As only 70 countries were represented in the consortium, we can test the effect of the shock in countries where

---

[31]The detailed methodology is available on the organization's website: `https://rsf.org/en/detailed-methodology`. The first group (*Satisfactory*) corresponds to countries with a score between 0 and 25, *Intermediate* ranges between 25 and 55, *Poor* is above 55.

[32]The Media Ownership and Concentration Diversity Index captures the degree of concentration of the market through the Herfindahl-Hirschman Index corrected for the number of players on the market, see Figure C.6 for the mathematical definition.

no media outlet was involved in the investigation process. Among these countries, the perception of the quality of journalism should not evolve differently following the revelation of the Panama Papers.

Hence, to compare the two groups of countries, we depart from Regression 1 and interact the treatment variable *Exposure* with a categorical variable characterizing the presence of at least one national outlet in the consortium. We can therefore distinguish between countries where the national media directly reported on the Panama Papers and those that were hit by the shock through foreign investigation.

We start with estimating the impact of the revelation on perceived corruption in Table D.14 and find that the shock has a similar effect in the two groups. We turn to media capture in Table 1.5. In countries out of the consortium, we observe a significant increase in internal capture of similar magnitude to what we find in countries represented in the consortium[33]. We conclude that our findings are valid in the absence of information about the quality of investigative journalism.

Demand for news

The Panama Papers could also have increased the demand for news in exposed countries. This would make the media market more profitable and potentially attract new owners, which could have triggered the effect on internal capture. To test the impact of the shock on the profitability of the media sector, we use data on advertising revenues and newspaper circulation from the World Association of News Publishers (WAN-IFRA). In Table D.15, we estimate a difference-in-differences specification similar to Regression 1.1, with advertising revenues and circulation of newspapers as dependent variables. Data on advertising are available for 57 countries and circulation for only 17 countries. All of them were affected by the Panama Papers, therefore, we cannot estimate the effect on the extensive margin. All our point estimates are non significant and very close to 0, which confirms that the exposure to the Panama Papers did not affect the demand for news.

Further remarks

Before we conclude, we would like to acknowledge two additional alternative mechanisms. Firstly, it could be that the share of owners with other economic interests increased because media owners diversified their activities. Since the profitability of the media sector does not seem affected, media owners could have seized this opportunity to invest in other sectors

---

[33]The coefficients are not significantly different from each other: at the intensive margin, the F-test p-value is .26.

and, in particular, in firms negatively affected by the leaks. Our data do not allow us to formally exclude this explanation. However, if it was the mechanism at play, the effect would be driven by countries where the leaks triggered important changes in the political and economic landscape. In particular, a change in government generally decreases the value of firms connected to the incumbent (Faccio 2006). To alleviate this concern, we estimate Regression 1.1 excluding countries with political turnover in the months after the scandal and find a very similar effect[34].

Secondly, the Panama Papers could have directly revealed to the principal that media outlets are capable to hurt his economic interests, without increasing media outlets' investigation. The effect would be larger in exposed countries and justify an increase in internal capture. This explanation is arguably a simplification of our model: internal capture increases due to a change in principal's beliefs, but this change is not justified by a strategic reaction of media outlets. This argument has similar implications for media capture and, we believe, does not challenge our conclusions.

5 Conclusion

In this paper, we argue that the different forms of media capture are not mere substitutes. We propose a theory where the forms of capture differ with respect to the control of journalistic investigation. Internal media capture, which includes, for instance, media ownership, involves early offers and prevents journalists from investigating sensitive topics. Instead, the principal can let media outlets investigate and potentially make ex-post offers to influence the content of publication. We predict that the likelihood of internal capture is increasing in perceived corruption: media outlets investigate more when they think the principal is corrupt, which increases the incentive to capture the media at an early stage. With respect to external capture, we expect a negative effect of perceived corruption if the media market is small enough.

We use the revelation of the Panama Papers to test these predictions. We first show that it is an exogenous shock to perceived corruption: after the revelation, perceptions increase more in countries that were exposed to the shock. Moreover, this change triggers an increase in internal media capture as well as a decrease in external capture in countries where the media market is less developed, which is in line with the prediction of the model.

Our paper underlines the necessity of a comprehensive approach of media capture. How-

---

[34]Between April 2016 and December 2017, government changed in 18% of the countries of the sample. Estimating Regression 1.1 in countries without political turnover, the effect on the share of media owners with outside economic interests is 0.935 (s.d. 0.389).

Table 1.5: The Panama Papers in countries in and out of the consortium

	(1) Owners	(2) Conflicts	(3) In-kind	(4) Bribes	(5) Owners	(6) Conflicts	(7) In-kind	(8) Bribes
After × Exposure								
In the consortium	1.248***	0.405	-0.136	-0.017				
	(0.469)	(0.473)	(0.193)	(0.193)				
Out of the consortium	0.879**	0.563	-0.180	0.048				
	(0.405)	(0.455)	(0.162)	(0.177)				
After × Exposure (log)								
In the consortium					0.108**	0.023	0.003	0.039*
					(0.053)	(0.047)	(0.021)	(0.021)
Out of the consortium					0.218***	0.049	0.039	0.024
					(0.068)	(0.055)	(0.041)	(0.038)
Observations	950	950	938	950	950	950	938	950
Mean DepVar	5.29	6.84	4.50	4.07	5.29	6.84	4.50	4.07
Sd DepVar	2.25	2.19	0.91	1.17	2.25	2.19	0.91	1.17
Adj. R2	0.53	0.60	0.49	0.65	0.53	0.60	0.49	0.65
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
p-val: In = Out	.26	.53	.77	.64	.10	.59	.35	.69

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimation of Regression 1.1, interactions with participation in ICIJ added. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. *In (resp. out of) the consortium*: Treatment variable is interacted with a variable equal to 1 if at least one national media outlet (resp. no outlet) was involved in the ICIJ. Dependent variables: *Owners*: proportion of general-interest media owned by companies with interests in other economic sectors. *Conflicts*: likelihood of journalist self-censorship due to conflicts of interest with the owner. *In-kind*: non-monetary transfers to journalists to influence the coverage. *Bribes*: payments to journalists to influence the coverage.

ever, deriving policy recommendations would require to quantify empirically the implications of the documented change in media capture strategy for editorial content. This is outside the scope of the current paper, but a growing literature attempts to measure these effects: Fize (2020) estimates the consequences of changes in ownership on the coverage of topics sensitive to the owner and Durante et al. (2020) show that bank lending affected the coverage of the Eurozone Sovereign Debt Crisis.Nevertheless, these findings are hard to reconcile in a single framework and further research is needed to quantify the implications of the change in capture strategy on media bias.

Finally, this paper is a first step in understanding how the forms taken by media capture depend on the environment. It is based on perceived corruption as a key determinant of media capture, but other parameters, such as media competition, also impacts the principal's trade-off and could affect capture strategies. We believe that a better understanding of such mechanisms is needed to identify the threats to media independence and implement regulations to effectively protect the freedom of information producers.

References

Alstadsæter, A., Johannesen, N., and Zucman, G. (2019). Tax Evasion and Inequality. *American Economic Review*, 109(6):2073–2103.

Ananyev, M., Xefteris, D., Zudenkova, G., and Petrova, M. (2019). Information and communication technologies, protests, and censorship. *SSRN Working Paper*.

Angelucci, C., Cagé, J., and Sinkinson, M. (2020). Media competition and news diets.

Beattie, G. (2020). Advertising and media capture: The case of climate change. *Journal of Public Economics*, 188:104219.

Beattie, G., Durante, R., Knight, B., and Sen, A. (2020). Advertising spending and media bias: Evidence from news coverage of car safety recalls. *Management Science*.

Besley, T. and Prat, A. (2006). Handcuffs for the grabbing hand? media capture and government accountability. *American Economic Review*, 96(3):720–736.

Bignon, V. and Flandreau, M. (2011). The economics of badmouthing: Libel law and the underworld of the financial press in france before world war i. *The Journal of Economic History*, 71(3):616–653.

Bignon, V. and Miscio, A. (2010). Media bias in financial newspapers: evidence from early twentieth-century france. *European Review of Economic History*, 14(3):383–432.

Brunetti, A. and Weder, B. (2003). A free press is bad news for corruption. *Journal of Public Economics*, 87(7-8):1801–1824.

Cagé, J. (2019). Media competition, information provision and political participation: Evidence from french local newspapers and elections, 1944–2014. *Journal of Public Economics*, page 104077.

Chowdhury, S. K. (2004). The effect of democracy and press freedom on corruption: an empirical test. *Economics letters*, 85(1):93–101.

Corneo, G. (2006). Media capture in a democracy: The role of wealth concentration. *Journal of Public Economics*, 90(1-2):37–58.

DellaVigna, S. and Hermle, J. (2017). Does conflict of interest lead to biased coverage? evidence from movie reviews. *Review of Economic Studies*, 84(4):1510–1550.

Di Tella, R. and Franceschelli, I. (2011). Government Advertising and Media Coverage of Corruption Scandals. *American Economic Journal: Applied Economics*, 3(4):119–151.

Djankov, S., McLeish, C., Nenova, T., and Shleifer, A. (2003). Who Owns the Media? *Journal of Law and Economics*, 46(2):341–381.

Durante, R. and Knight, B. (2012). Partisan control, media bias, and viewer responses: Evidence from berlusconi's italy. *Journal of the European Economic Association*, 10(3):451–481.

Durante, R., Peydró, J.-L., et al. (2020). Media capture by banks.

Dyck, A., Volchkova, N., and Zingales, L. (2008). The corporate governance role of the media: Evidence from russia. *The Journal of Finance*, 63(3):1093–1135.

Dziuda, W. and Howell, W. G. (2021). Political scandal: a theory. *American Journal of Political Science*, 65(1):197–209.

Edmond, C. (2013). Information manipulation, coordination, and regime change. *Review of Economic Studies*, 80(4):1422–1458.

Enikolopov, R. and Petrova, M. (2015). Media capture: empirical evidence. In *Handbook of media Economics*, volume 1, pages 687–700. Elsevier.

Eraslan, H. and Özertürk, S. (2018). Information gatekeeping and media bias. Technical report, working paper.

Faccio, M. (2006). Politically connected firms. *American economic review*, 96(1):369–386.

Fize, E. (2020). Acquiring newspapers, acquiring influence? media bias and capture by billionaires. *Unpublished*.

Freille, S., Haque, M. E., and Kneller, R. (2007). A contribution to the empirics of press freedom and corruption. *European Journal of Political Economy*, 23(4):838–862.

Furceri, D., Papageorgiou, C., and Ahir, H. (2019). Global incidents of corruption index. *IMF*.

Gehlbach, S. and Sonin, K. (2014). Government Control of the Media. *Journal of Public Economics*, 118:163–171.

Gentzkow, M. and Shapiro, J. M. (2006). Media Bias and Reputation. *Journal of Political Economy*, 114(2).

Germano, F. and Meier, M. (2013). Concentration and self-censorship in commercial media. *Journal of Public Economics*, 97:117–130.

Gilens, M. and Hertzman, C. (2000). Corporate ownership and news bias: Newspaper coverage of the 1996 telecommunications act. *The Journal of Politics*, 62(2):369–386.

Gitmez, A. and Molavi, P. (2018). Media capture: a bayesian persuasion approach.

Gratton, G. (2015). The sound of silence: Political accountability and libel law. *European Journal of Political Economy*, 37:266–279.

Guriev, S., Melnikov, N., and Zhuravskaya, E. (2019). 3g internet and confidence in government. *SSRN Working Paper*.

Guriev, S. and Treisman, D. (2019). Informational autocrats. *Journal of Economic Perspectives*, 33(4):100–127.

Kaufmann, D., Kraay, A., and Mastruzzi, M. (2011). The worldwide governance indicators: methodology and analytical issues. *Hague Journal on the Rule of Law*, 3(2):220–246.

Mastrorocco, N. and Ornaghi, A. (2020). Who watches the watchmen? local news and police behavior in the united states.

McMillan, J. and Zoido, P. (2004). How to Subvert Democracy: Montesinos in Peru. *Journal of Economic Perspectives*, 18(4):69–92.

Miho, A. (2020).

Mullainathan, S. and Shleifer, A. (2005). The market for news. *American Economic Review*, 95(4):1031–1053.

Noam, E. M. (2016). *Who owns the world's media?: media concentration and ownership around the world*. Oxford University Press.

O'Donovan, J., Wagner, H. F., and Zeume, S. (2019). The value of offshore secrets: evidence from the panama papers. *The Review of Financial Studies*, 32(11):4117–4155.

Olken, B. A. (2009). Corruption perceptions vs. corruption reality. *Journal of Public Economics*, 93(7-8):950–964.

Petrova, M. (2008). Inequality and media capture. *Journal of Public Economics*, 92(1-2):183–212.

Prat, A. and Strömberg, D. (2013). The political economy of mass media. *Advances in economics and econometrics*, 2:135.

Reporters Without Borders (2016). Media oligarchs go shopping. Technical report.

Reuter, J. and Zitzewitz, E. (2006). Do ads influence editors? advertising and bias in the financial media. *The Quarterly Journal of Economics*, 121(1):197–227.

Rizzica, L. and Tonello, M. (2015). Exposure to media and corruption perceptions. *Bank of Italy Temi di Discussione (Working Paper) No*, 1043.

Shadmehr, M. and Bernhardt, D. (2015). State censorship. *American Economic Journal: Microeconomics*, 7(2):280–307.

Snyder, J. J. and Strömberg, D. (2010). Press coverage and political accountability. *Journal of Political Economy*, 118(2):355–408.

Stanig, P. (2015). Regulation of speech and media coverage of corruption: An empirical analysis of the mexican press. *American Journal of Political Science*, 59(1):175–193.

Stiglitz, J. (2017). Toward a Taxonomy of Media Capture. In Schiffrin, A., editor, *In the Service of Power: Media Capture and the Threat to Democracy*, pages 9–17. Center for International Media Assistance.

Szeidl, A. and Szucs, F. (2021). Media capture through favor exchange. *Econometrica*, 89(1):281–310.

Transparency International (2018). Corruption Perceptions Index 2018: Technical Methodology Note. Technical report.

Zucman, G. (2013). The missing wealth of nations: Are europe and the us net debtors or net creditors? *The Quarterly Journal of Economics*, 128(3):1321–1364.

Appendices

A   Proofs

Lemma 1

Proof. Suppose $q_1^* \geq ... \geq q_n^*$ is an equilibrium of the investigation stage. We consider the media outlets in turn:

- $i > s$. As long as $q_{i-1} \geq q_i \geq q_{i+1}$ (noting $q_{n+1} = 0$), i's expected profit is,

$$\pi(q_i) = P[\theta = c|t^I]a\Big[\sum_{j=0}^{n-i-1} \frac{q_{n-j} - q_{n-j+1}}{n-j} + \frac{q_i - q_{i+1}}{i} \Big] - c(q_i),$$

  The F.O.C. gives:
$$q_n^* = c'^{-1}(\frac{P[\theta = c|t^I]a}{i}).$$

- $s \leq i$. For $q_i \geq q_{s+1}$,

$$\pi(q_i) = P[\theta = c|t^I]a\Big[\sum_{j=0}^{n-s-1} \frac{q_{n-j} - q_{n-j+1}}{n-j} + q_i - q_{s+1} \Big] - c(q_i),$$

  and the F.O.C. implies:
$$q_i^* = c'^{-1}(P[\theta = c|t^I]a).$$

Finally, to see that the F.O.C. are sufficient to guarantee that there is no profitable deviation, notice that for all i, $\pi'(q_i) > 0$ for $q_i < q_i^*$ and $\pi'(q_i) < 0$ for $q_i > q_i^*$.

$\square$

Lemma 2

Proof. We first show that the set of offers t^I that are not weakly dominated for all parameters consists in three elements: $\{(t_i^I = 0 \ \forall i), (t_i^I = \pi_1^{\theta=c} \ \forall i), (t_i^I = 0 \text{ for } i \leq s, t_i^I = \pi_{s+1}^{\theta=c} \text{ for } i > s)\}$, with:
$$\pi_1^{\theta=c} = \max_{q_i}\{aq_i - c(q_i)\}$$
and
$$\pi_{s+1}^{\theta=c} = \max_{q_i}\{\frac{a}{s+1}q_i - c(q_i)\}.$$

- Equilibrium offers match exactly the deviation payoff of media outlets: smaller offers would be rejected while larger offers do not minimize cost.
- Capturing less than $n - s$ media outlets does not affect the principal's continuation payoff and cannot be an equilibrium strategy.
- If media outlets $i > s + 1$ are captured, each media outlet $1 < i < s + 1$ costs $\pi_1^{\theta=c}$ in internal capture and $q_1^{\theta=c} \times a$, where $q_1^{\theta=c} = \arg\max_{q_i}\{aq_i - c(q_i)\}$ in expectation if it is not internally captured. Thus, the principal either captures all or none of them[35].
- Media outlets' payoffs are decreasing in i, the principal captures the outlets with the largest indexes.

Only the 3 offers proposed above satisfy all conditions. Moreover, they yield the following expected costs for the principal:

$$t_i^I = 0 \ \forall i \rightarrow s \times a \times (q_s^* - q_{s+1}^*) + r q_{s+1}^*,$$
$$t_i^I = \pi_1^{\theta=c} \ \forall i \rightarrow n \times \pi_1^{\theta=c},$$
$$(t_i^I = 0 \text{ for } i \leq s, t_i^I = \pi_{s+1}^{\theta=c} \text{ for } i > s) \rightarrow (n-s)\pi_{s+1}^{\theta=c} + q_1^{\theta=c} \times a \times s,$$

where $q_1^{\theta=c} = \arg\max_{q_i}\{aq_i - c(q_i)\}$. As a result, we have[36]:

$$n_0 = \frac{s[aq_1^{\theta=c} - \pi_{s+1}^{\theta=c}]}{\pi_1^{\theta=c} - \pi_{s+1}^{\theta=c}}$$

and

$$n_1 = \max\{\frac{s(a(q_s^* - q_{s+1}^* - q_1^{\theta=c}) + \pi_{s+1}^{\theta=c}) + r q_{s+1}^*}{\pi_{s+1}^{\theta=c}}, \frac{sa(q_s^* - q_{s+1}^*) + r q_{s+1}^*}{\pi_1^{\theta=c}}\}.$$

$\square$

Proposition 1

Proof. The principal plays $t_i^I = 0 \ \forall i$ with probability $1 - x$ and the cheapest element of $\{(t_i^I = \pi_1^{\theta=c} \ \forall i), (t_i^I = 0 \text{ for } i \leq s, t_i^I = \pi_{s+1}^{\theta=c} \text{ for } i > s)\}$ with probability x. We write C^I the cost of the second option. It does not depend on x and γ. The cost of the first option is $C^E(x, \gamma)$:

$$C^E(x, \gamma) = a \times s[q_1^* - q_{s+1}^*] + r q_{s+1}^*.$$

It is decreasing in x and increasing in γ, as explained in the text.

We have three possible cases: (i) $C^I > C^E(0, \gamma)$ and $x^* = 0$, (ii) $C^I < C^E(1, \gamma)$ and $x^* = 1$ or (iii) $C^E(1, \gamma) \leq C^I \leq C^E(0, \gamma)$ and x^* satisfies

$$C^E(x^*(\gamma, n), \gamma) = C^I. \tag{1.2}$$

---

[35]If $\pi_1^{\theta=c} = q_1^{\theta=c} \times a$, the principal is exactly indifferent and could buy $n - s < k < n$ media outlets. Unless k depends on the probability of internal capture x, other results would not be affected.

[36]Special cases: as we assume $n > s + 1$, we can have:

- $n_1 < s + 1$: for all n, the principal chooses $t_i^I = 0 \ \forall i$,
- $n_1 < n_0$: there is no n such that the principal chooses $(t_i^I = 0 \text{ for } i \leq s, t_i^I = \pi_{s+1}^{\theta=c} \text{ for } i > s)$,
- $n_0 < s + 1$: there is no n such that the principal chooses $t_i^I = \pi_1^{\theta=c} \ \forall i$.

When x^* is defined by (1.2), we have:

$$\frac{\partial x^*(\gamma, n)}{\partial \gamma} = -\frac{\partial C^E(x^*, \gamma)/\partial \gamma}{\partial C^E(x^*, \gamma)/\partial x^*} > 0.$$

□

Corollary 1

Proof. $C^I(n)$ is increasing in n while $C^E(x, \gamma)$ is independent of n. Thus, we have:

$$\frac{\partial x^*(\gamma, n)}{\partial n} = \frac{\partial C^I(n)/\partial n}{\partial C^E(x^*, \gamma)/\partial x^*} < 0.$$

□

Proposition 2

Proof. We show that:

- For $n < n_0$, $\frac{\partial P^E(\gamma)}{\partial \gamma} < 0$,
- For $n_0 \le n \le n_1$, $\frac{\partial P^E(\gamma)}{\partial \gamma}$ is increasing in n with $\lim_{n \to n_1} \frac{\partial P^E(\gamma)}{\partial \gamma} > 0$,
- For $n > n_1$, $\frac{\partial P^E(\gamma)}{\partial \gamma} > 0$.

Thus, there must exist a $\bar{n}$ such that $P^E(\gamma)$ is decreasing in γ if $n < \bar{n}$ and increasing otherwise[37].

$\underline{n < n_0}$: Principal buys all outlets.

$$P^E(\gamma) = (1 - x^*(\gamma))(q_1^* - q_{s+1}^*)$$
$$= \frac{a\gamma s(1 - x^*(\gamma))^2}{\kappa(s + 1)(1 - \gamma x^*(\gamma))}$$
$$\Leftrightarrow \frac{\partial P^E(\gamma)}{\partial \gamma} = \frac{as(1 - x^*(\gamma))\left(\gamma(\gamma x^*(\gamma) + \gamma - 2)(\partial x(\gamma)/\partial \gamma) + 1 - x^*(\gamma)\right)}{\kappa(s + 1)(\gamma x^*(\gamma) - 1)^2},$$

where $x^*(\gamma)$ solves

$$n\left(\frac{a^2}{\kappa} - c(\frac{a}{\kappa})\right) = \frac{a\gamma(1 - x)\left(as^2 + r\right)}{\kappa(s + 1)(1 - \gamma x)},$$

which implies:

$$x^*(\gamma) = \frac{a\left(a\gamma s^2 - a(s + 1)n + \gamma r\right) + \kappa(s + 1)nc(\frac{a}{\kappa})}{a\gamma\left(a\left(s^2 - (s + 1)n\right) + r\right) + \gamma\kappa(s + 1)nc(\frac{a}{\kappa})},$$
$$\frac{\partial x^*(\gamma)}{\partial \gamma} = \frac{(s + 1)n\left(a^2 - \kappa c(\frac{a}{\kappa})\right)}{\gamma^2\left(a\left(a\left(s^2 - (s + 1)n\right) + r\right) + \kappa(s + 1)nc(\frac{a}{\kappa})\right)}.$$

---

[37] The proof assumes $n_0 < n_1$. Otherwise, the second part of the proof can be omitted and the proposition holds with $\bar{n} = n_0$.

Notice that the sign of $\frac{\partial P^E(\gamma)}{\partial \gamma}$ uniquely depends on the sign of:

$$\gamma(\gamma x^*(\gamma) + \gamma - 2)(\partial x(\gamma)/\partial \gamma) + 1 - x^*(\gamma).$$

We plug the expressions of $x^*(\gamma)$ and $\frac{\partial x^*(\gamma)}{\partial \gamma}$ and obtain:

$$\frac{a(1-\gamma)(s+1)n\left(-a^2 + \kappa c(\frac{a}{\kappa})\right)(as^2 + r)}{\gamma\left(a\left(a\left(s^2 - (s+1)n\right) + r\right) + \kappa(s+1)nc(\frac{a}{\kappa})\right)^2},$$

the sign of which depends on

$$-a^2 + \kappa c(\frac{a}{\kappa}) = -\kappa \times \pi_1^{\theta = c} < 0.$$

Thus, if $n < n_0$, $\frac{\partial P^E(\gamma)}{\partial \gamma} < 0$.

<u>$n_0 \leq n \leq n_1$: Principal buys the $n - s$ last media outlets.</u>

$$P^E(\gamma) = x^*(\gamma)q_1^{\theta=c} + (1 - x^*(\gamma))(q_1^* - q_{s+1}^*)$$
$$= \frac{ax^*(\gamma)}{\kappa} + \frac{a\gamma s(1 - x^*(\gamma))^2}{\kappa(s+1)(1 - \gamma x^*(\gamma))}$$
$$\Leftrightarrow \frac{\partial P^E(\gamma)}{\partial \gamma} = \frac{a\left((\gamma-1)^2 s + \gamma x^*(\gamma)(\gamma x^*(\gamma) - 2) + 1\right)(\partial x(\gamma)/\partial \gamma) + as(x^*(\gamma) - 1)^2}{\kappa(s+1)(\gamma x^*(\gamma) - 1)^2}$$

$x^*(\gamma)$ solves:

$$\frac{a^2 s}{\kappa} + (n - s)\left(\frac{a^2}{\kappa(s+1)^2} - c\left(\frac{a}{\kappa(s+1)}\right)\right) = \frac{a\gamma(1-x)(as^2 + r)}{\kappa(s+1)(1 - \gamma x)}.$$

Which implies:

$$x^*(\gamma) = \frac{-a\left(as^2((\gamma-1)s + \gamma - 2) - an + \gamma(s+1)r\right) + \kappa(s+1)^2(s-n)c\left(\frac{a}{\kappa(s+1)}\right)}{\gamma\left(a\left(a\left(s^2 + n\right) - (s+1)r\right) + \kappa(s+1)^2(s-n)c\left(\frac{a}{\kappa(s+1)}\right)\right)}$$

$$\frac{\partial x^*(\gamma)}{\partial \gamma} = -\frac{a^2\left((s+2)s^2 + n\right) + \kappa(s+1)^2(s-n)c\left(\frac{a}{\kappa(s+1)}\right)}{\gamma^2\left(a\left(a\left(s^2 + n\right) - (s+1)r\right) + \kappa(s+1)^2(s-n)c\left(\frac{a}{\kappa(s+1)}\right)\right)}$$

Notice that the sign of $\frac{\partial P^E(\gamma)}{\partial \gamma}$ uniquely depends on the sign of:

$$a\left((\gamma-1)^2 s + \gamma x^*(\gamma)(\gamma x^*(\gamma) - 2) + 1\right)(\partial x(\gamma)/\partial \gamma) + as(x^*(\gamma) - 1)^2$$

72

and plug the expressions of $x^*(\gamma)$ and $\frac{\partial x^*(\gamma)}{\partial \gamma}$. We obtain:

$$\left\{ -a^2(\gamma - 1)^2(s + 1)\left(as^2 + r\right)\left(a^2\left((s + 2)s^2 + n\right) + \kappa(s + 1)^2(s - n)c\left(\frac{a}{\kappa(s + 1)}\right)\right)\right.$$

$$\left(a\left(as(s - n) + (s + 1)^2r\right) - \kappa s(s + 1)^2(s - n)c\left(\frac{a}{\kappa(s + 1)}\right)\right)\right\} \Big/$$

$$\left\{\gamma^2\left(a\left(a\left(s^2 + n\right) - (s + 1)r\right) + \kappa(s + 1)^2(s - n)c\left(\frac{a}{\kappa(s + 1)}\right)\right)^3\right\}.$$

Notice

$$-a^2(\gamma - 1)^2(s + 1)\left(as^2 + r\right) < 0,$$

moreover,

$$a^2\left((s + 2)s^2 + n\right) + \kappa(s + 1)^2(s - n)c\left(\frac{a}{\kappa(s + 1)}\right)$$

$$= n\kappa(s + 1)^2\pi_{s+1}^{\theta=c} + (s + 1)^2\kappa s + a^2(s + 2)s > 0.$$

We can focus on the sign of:

$$-\frac{\left(a\left(as(s - n) + (s + 1)^2r\right) - \kappa s(s + 1)^2(s - n)c\left(\frac{a}{\kappa(s+1)}\right)\right)}{\left(a\left(a\left(s^2 + n\right) - (s + 1)r\right) + \kappa(s + 1)^2(s - n)c\left(\frac{a}{\kappa(s+1)}\right)\right)^3} = A(n).$$

We take the derivative of this expression with respect to n:

$$\left(a(as(s(s + 3) - 2n) + (s + 1)(2s + 3)r) - 2\kappa s(s + 1)^2(s - n)c\left(\frac{a}{\kappa(s + 1)}\right)\right)$$

$$\times \frac{\left(k(s + 1)^2 \times \pi_{s+1}^{\theta=c}\right)}{\left(a(a(s^2 + n) - (s + 1)r) + \kappa(s + 1)^2(s - n)c\left(\frac{a}{\kappa(s+1)}\right)\right)^4}$$

the sign of which depends on:

$$a(as(s(3 + s) - 2n) + (1 + s)(3 + 2s)r) - 2\kappa s(1 + s)^2(s - n)c\left(\frac{a}{\kappa(s + 1)}\right)$$

$$= -n[2\kappa s(s + 1)^2\pi_{s+1}^{\theta=c}]$$

$$\quad + a(as(s(3 + s)) + (1 + s)(3 + 2s)r) - 2\kappa s^2(1 + s)^2c\left(\frac{a}{\kappa(s + 1)}\right)$$

$$= B(n)$$

This expression is decreasing in n. However, the case assumption requires $n < n_1$, with

$$n_1 = \frac{a\left(as^2((\gamma - 1)s + \gamma - 2) + \gamma(s + 1)r\right) - \kappa s(s + 1)^2c\left(\frac{a}{\kappa(s+1)}\right)}{a^2 - \kappa(s + 1)^2c\left(\frac{a}{\kappa(s+1)}\right)}.$$

Plugging this expression in $B(n)$, we obtain:

$$B(n_1) = a(1+s)(3+2(1-\gamma)s)(as^2+r) > 0.$$

Thus, for $n_0 \leq n \leq n_1$, $B(n) > 0$ and $\frac{\partial A(n)}{\partial n} > 0$. Finally:

$$A(n_1) = \frac{1+(1-g)s}{a^2(1-g)^3(s+1)^2 \left(as^2+r\right)^2} > 0.$$

As a result, for $n_0 \leq n \leq n_1$, $\frac{\partial P^E(\gamma)}{\partial \gamma}$ is increasing in n with $\lim_{n \to n_1} \frac{\partial P^E(\gamma)}{\partial \gamma} > 0$.

$\underline{n > n_1}$: No internal capture.

$$\begin{aligned}
P^E(\gamma) &= q_1^* - q_{s+1}^* \\
&= \frac{a}{\kappa}\frac{s}{s+1}\gamma \\
\Leftrightarrow \frac{\partial P^E(\gamma)}{\partial \gamma} &= \frac{a}{\kappa}\frac{s}{s+1} > 0.
\end{aligned}$$

$\square$

B Alternative modelling assumptions

B.1 Endogenous corruption

In this extension, we are interested in the effect of actual corruption on media capture. We consider a setup where the principal endogenously choses his type at the beginning of the game. His payoff of choosing $\theta = h$ is normalized to 0 while $\theta = c$ provides a positive payoff β and implies to pay the costs of capture or punishment r discussed in the main text.

The next proposition shows that our main predictions on the perception of corruption, namely Propositions 1 and 2, as well as Corollary 1, also apply to the reward of corruption β. One difference should however be noticed: when the number of media outlets is small, the probability of external capture is constant in β (while it decreases with the prior in the main model).

Proposition 3. *1. The probability of internal capture is increasing in the payoff from corruption β.*

2. The probability of internal capture is decreasing in the number of media outlets n.

3. Under Assumption A, there exists a number of media outlets $\underline{n}$ such that the probability of external capture is increasing in β if $n > \underline{n}$ and does not depend on n otherwise.

Proof. 1. The probability of internal capture is increasing in β.

We note γ the probability that the principal engages in corruption. External capture, investigation and internal capture stages are similar to the main model. There exists $\underline{\gamma}$ such

that, as long as $\gamma < \underline{\gamma}$, the principal never uses internal capture ($x^* = 0$). If $\gamma > \underline{\gamma}$, the probability of internal capture x^* equalizes the cost of internal and external capture.

Turning to the first stage, where the principal chooses whether to engage in corruption, the trade-off is:

$$\text{Corrupt} \rightarrow \beta - \text{Cost of capture}(\gamma)$$
$$\text{Honest} \rightarrow 0.$$

For $\gamma < \underline{\gamma}$, the cost of capture is increasing in γ. For $\gamma \geq \underline{\gamma}$, it is equal to the cost of internal capture and, therefore, does not depend on γ. We define $\underline{\beta}$ such that $\underline{\beta} = \text{Cost of capture}(\underline{\gamma})$. We have:

- For $\beta < \underline{\beta}$, $\frac{\partial \gamma^*(\beta)}{\partial \beta} > 0$ and $x^* = 0$,
- for $\beta \geq \underline{\beta}$, $\gamma^* = 1$ and $x^* = 1$.

The probability of internal capture is:

$$P^I(\beta) = x^* \gamma^* = \begin{cases} 0 & \text{if } \beta < \underline{\beta}, \\ 1 & \text{if } \beta \geq \underline{\beta}. \end{cases}$$

Thus, P^I is discontinuously increasing in β.

2. The probability of internal capture is decreasing in n.
First of all, we show that $\underline{\beta}$ is increasing in n. $\underline{\beta}$ solves:

$$C^I = C^E(\gamma^*(\underline{\beta}), x^* = 0).$$

The r.h.s. is increasing in β, as γ^* is increasing in β and C^E is increasing in γ. Moreover, the r.h.s. is independent of n while the l.h.s. is increasing in n. Thus, $\underline{\beta}$ is increasing in n.

We can now conclude that P^I decreases in n from its expression: when n increases, the probability of internal capture decreases from 1 to 0 for $\beta \rightarrow \underline{\beta}$ and does not vary otherwise.

3. There exists $\underline{n}(\beta)$ such that P^E is increasing in γ if and only if $n > \underline{n}(\beta)$.
If $\beta < \underline{\beta}$, $x^* = 0$. Otherwise, $x^* = 1$. Thus, if $\beta < \underline{\beta}$, the probability of external capture is

$$P^E(\beta) = \gamma^*[q_1^* - q_{s+1}^*].$$

When $\beta \geq \underline{\beta}$, we have:

$$P^E(\beta) = \begin{cases} 0 & \text{if } n < n_0, \\ q_s & \text{if } n_0 < n < n_1. \end{cases}$$

(Notice that $n > n_1$ implies $x^* = 0$, and thus $\beta < \underline{\beta}$).

- If $\beta < \underline{\beta}$, we have established that γ^* increases in β. Under Assumption A and given $x^* = 0$, $q_1^* - q_{s+1}^* = \gamma \frac{s}{s+1} a$. Thus, P^E is increasing in β.
- If $\beta \geq \underline{\beta}$, P^E does not depend on β.
- As established in 2., $\underline{\beta}$ is increasing in n. Thus, for a given β, P^E is increasing in β if n is larger than some $\underline{n}$ and constant otherwise.

$\square$

B.2   Independent signals

We consider an alternative setup where media outlets receive their signal independently from each other. We focus on $n \in \{1, 2\}$ and assume $a < r < 2a$. Payoffs, timing and information structure are similar to the main model.

The next proposition shows that the main predictions of our model are not affected: the probability of internal capture is increasing in the prior γ. Moreover, as established in Corollary 1, internal capture is less likely when the media market is competitive. Finally, if the cost of investigation is quadratic, the probability of external capture can increase with the prior if $n = 1$ but decrease if $n = 2$.

Proposition 4. *When signals are independent,*
1. *the probability of internal capture x^* is increasing in γ,*
2. *x^* is larger if $n = 1$ than if $n = 2$,*
3. *assume $c(q) = q^2/2$. Then, if for $n = 1$ the probability of external capture P^E increases in γ, then it also increases for $n = 2$. However, P^E can be increasing for $n = 2$ but decreasing for $n = 1$.*

Proof. 1. x^* is increasing in γ.

- $\underline{n = 1.}$ In the external capture stage, the principal offers $t^E = a$ if the media outlet has found the signal and suppresses it. In the investigation stage, the media outlet maximizes:
$$\pi(q) = aP[\theta = c|t^I]q - c(q),$$
and the F.O.C. gives:
$$c'(q^*) = aP[\theta = c|t^I].$$
As in the main text, q^* is decreasing in the probability of internal capture x and increasing in the prior γ. If the principal chooses $t^I = 0$, the expected cost is aq^*.
In the internal capture stage, the principal offers the deviation payoff of the media outlet:
$$\pi^{\theta=c} = aq^{\theta=c} - c(q^{\theta=c}),$$
where $q^{\theta=c} = \arg\max_q\{aq - c(q)\}$. This term does not depend on x and γ. The equilibrium probability of internal capture x^* equalizes the expected costs of internal and external capture and an inspection of the two functions completes the proof.
- $\underline{n = 2.}$ In the external capture stage, the principal suppresses the signal if one media outlet receives it and lets media outlets publish if both of them find it. The expected payoff of media outlet i is:
$$\pi_i(q_i) = a(1 - \frac{q_j}{2})P[\theta = c|t^I]q_i - c(q_i),$$

76

and the equilibrium level of investigation of both outlets solves:

$$c'(q^{**}) = a(1 - \frac{q^{**}}{2})P[\theta = c|t^I]$$

Consider internal capture. $t^I = \{0, 0\}$ yields an expected cost

$$C^E(x, \gamma) = a \times 2q^{**}(1 - q^{**}) + r \times (q^{**})^2.$$

$C^E(x, \gamma)$ is increasing in γ and decreasing in x. Turning to the cost of internal capture, the principal can buy one or two outlets. In the first case, he offers $t^I = \{\pi_2^{\theta=c}, 0\}$, where $\pi_2^{\theta=c}$ is the expected payoff on the media market with 2 free media outlets and $\theta = c$. The expected cost for the principal is $\pi_2^{\theta=c} + aq^{\theta=c}$. Buying the two outlets implies $t^I = \{\pi^{\theta=c}, \pi^{\theta=c}\}$ and yields an expected cost $2\pi^{\theta=c}$. The principal chooses the cheapest of the two options and the number of media outlets bought depends on the parameters. In both cases, the cost of internal capture is independent of x and γ. Thus, the proof of Proposition 1.1 applies and x^* is increasing in γ.

2. x^* is larger if $n = 1$ than if $n = 2$. Suppose $0 < x^* < 1$ (the proof generalizes for corner solutions). If $n = 1$, after rearranging, x^* satisfies:

$$2aq^* = 2\pi^{\theta=c},$$

while for $n = 2$:

$$2aq^{**} + (q^{**})^2(r - 2a) = \min\{2\pi^{\theta=c}, \pi_2^{\theta=c} + aq^{\theta=c}\}.$$

In both cases, the r.h.s. does not depend on x and is smaller for $n = 2$. Moreover, the l.h.s. is decreasing in x and is (weakly) smaller for $n = 2$. Thus, the equilibrium probability of internal capture is necessarily smaller for $n = 2$.

3. If the probability of external capture P^E increases in γ for $n = 1$, then it also increases for $n = 2$. However, P^E can be increasing for $n = 2$ but decreasing for $n = 1$.

With $c(q) = \frac{q^2}{2}$, we have for $n = 1$: $q^* = a\frac{\gamma(1-x)}{\gamma(1-x)+1-\gamma}$ (we set $a \leq 1$ to satisfy interiority condition), $C^I = \frac{a^2}{2}$ and $C^E = aq^*$, implying $x^* = (2\gamma - 1)/\gamma$ if $\gamma > 1/2$ and $x^* = 0$ otherwise. As a result:

$$P_{n=1}^E = \{ \begin{array}{ll} a\gamma & \text{if } \gamma < \frac{1}{2} \\ \frac{1-\gamma}{\gamma} \times \frac{a}{2} & \text{if } \gamma \geq \frac{1}{2}. \end{array}$$

Thus, when $n = 1$, the probability of external capture is increasing in γ if and only if $\gamma < \frac{1}{2}$.

For $n = 2$, we have shown that $x^*(\gamma)$ is increasing in γ. Moreover, there exists $\underline{\gamma}$ such that $x^*(\gamma) = 0$ iff $\gamma \leq \underline{\gamma}$ and we have:

$$P_{n=2}^E = \{ \begin{array}{ll} 2q^{**}(1 - q^{**}) & \text{if } \gamma < \underline{\gamma} \\ (1 - x^*)2q^{**}(1 - q^{**}) & \text{if } \gamma \geq \underline{\gamma}. \end{array}$$

With this cost function, notice that the principal prefers to buy 2 rather than 1 media outlet in internal capture. Buying 2 outlets costs a^2. If only one outlet is bought, the free outlet

chooses $q = a$ and the principal pays in expectation a^2 plus the cost of the internally captured media.

Each media outlet chooses:

$$q^{**} = \frac{2a\gamma(x-1)}{a\gamma(x-1) + 2\gamma x - 2},$$

which is $2a\gamma/(2 + a\gamma)$ for $x = 0$. Thus, for $\gamma < \underline{\gamma}$,

$$P^E_{n=2} = \frac{4a\gamma(2 - a\gamma)}{(a\gamma + 2)^2}.$$

Taking the derivative, we get:

$$\frac{8a(2 - 3a\gamma)}{(a\gamma + 2)^3},$$

which has the sign of $2 - 3a\gamma$. As $a \leq 1$, this is necessarily positive for $\gamma < 2/3$.

Now, we show that $\underline{\gamma} > 1/2$. For the principal, the equilibrium probability of internal capture solves $a^2 = a \times 2q^{**}(1 - q^{**}) + r(q^{**})^2$ and we have:

$$x^* = \frac{\gamma^2(a(a+6) - 4(r+1)) + 4\sqrt{-(\gamma-1)^2 g^2(2a - r - 1)} + 2a\gamma}{\gamma^2(a(a+8) - 4(r+1))}.$$

Setting this expression to 0, we obtain:

$$\underline{\gamma} = \frac{1}{\sqrt{-2a + r + 1 - \frac{a}{2} + 1}},$$

which is increasing in a and decreasing in r. However, as $r < 2a$ and $a \leq 1$, we must have $\underline{\gamma} > 1/2$.

As a result, we have established that $P^E_{n=2}$ is necessarily increasing in γ for all γ smaller than $\min\{\underline{\gamma}, 2/3\} > 1/2$. As $P^E_{n=1}$ is increasing in γ if and only if $\gamma < 1/2$, we have proved the claim.

$\square$

B.3 Bayesian persuasion

In our main model, a signal needs to be published for the principal to be punished. However, if the prior γ is large and the media never publish, the absence of a signal might not be sufficient to convince citizens not to punish the principal. In this appendix, we consider a setup where the punishment of the principal depends on the ex-post belief of a third party, which we call citizens, who observe the report of a unique media outlet.

We adopt a Bayesian persuasion approach and let the principal commit to a capture scheme (x^I, x^E) at the beginning of the game, where x^I is the probability of making an internal capture offer and x^E the probability of making an ex-post offer, conditional on the media outlet receiving a signal[38]. We posit a quadratic cost of investigation: $c(q) = q^2/2$.

---

[38] As in Gehlbach and Sonin (2014), this approach solves the commitment problem of the principal.

At the end of the game, citizens observe the report of the media $\sigma \in \{\text{Corrupt}, \emptyset\}$, as well as the capture scheme (x^I, x^E). They form an ex-post belief about the type of the principal and punish him if this posterior exceeds a critical value $\tau < 1$, that is, if $P(\theta = c | \sigma, (x^I, x^E), \gamma) > \tau$. τ is common knowledge. As in the main model, the principal's punishment cost is larger than the media outlet's publication payoff: $r > a$.

We solve the game backward and first consider the citizens. If the media outlet reports that the principal is corrupt, they punish him. Conditional on not receiving a signal, citizens use Bayesian updating to infer the probability that the principal is corrupt:

$$P(\theta = c | \sigma = \emptyset, (x^I, x^E), \gamma) = \frac{\gamma((1 - x^I)(q^* x^E - q^* + 1) + x^I)}{\gamma((1 - x^I)(q^* x^E - q^* + 1) + x^I) - \gamma + 1},$$

where q^* is the media outlet's investigation effort, which we now determine. If it has not been captured, the media outlet uses the probability of internal capture x^I to update its belief about the type of the principal. Given that the investigation cost is quadratic, it maximizes:

$$\pi(q) = P(\theta = c | t^I = 0, x^I) - \frac{q^2}{2}.$$

The F.O.C. combined with Bayes' law implies:

$$q^* = \frac{a\gamma(1 - x^I)}{\gamma(1 - x^I) + 1 - \gamma}.$$

We turn to the problem of the principal, who commits to probabilities of media capture (x^I, x^E). As in the main text, the principal's offers match the media outlet's deviation payoff: $t^I = \frac{a^2}{2} = \pi_1$ and $t^E = a$.

Key in this setup is the credibility of media capture. Conditional on not receiving a signal, citizens' posterior belief needs to be smaller than τ; otherwise, the principal is always punished and media capture is useless. If the prior γ is lower than the posterior punishment threshold τ, the credibility constraint never binds: the principal is not punished even if the signal is never disclosed. In such a case, the tradeoff is similar to the main model: if a signal is received, $a < r$ guarantees that the principal always suppresses it ex-post ($x^E = 1$) and the probability of internal capture x^I is increasing in γ because an early transfer is more profitable when the media outlet chooses a high investigation effort.

We now focus on the case $\gamma > \tau$. The principal must commit to let his type be revealed with some probability to make his intervention credible. As in Kamenica and Gentzkow (2011), the optimal capture scheme (x^I, x^E) is such that citizens' ex-post belief is exactly τ: if $P(\theta = c | \sigma = \emptyset, (x^I, x^E), \gamma) < \tau$, the principal could increase media capture and still be credible while if $P(\theta = c | \sigma = \emptyset, (x^I, x^E), \gamma) > \tau$, the principal is punished regardless of the signal received by citizens and should reduce capture.

Thus, we must have $P(\theta = c | \sigma = \emptyset, (x^I, x^E), \gamma) = \tau$, which can be used to express x^E as a function of x^I:

$$x^E = 1 - \frac{\gamma - \tau}{(1 - x^I)(1 - \tau)\gamma q^*}.$$

x^E is decreasing in x^I: when the principal increases internal capture, he must decrease ex-

ternal capture to keep the credibility constraint satisfied. Moreover, there is also a positivity constraint, as the credibility constraint must be satisfied for non-negative x^E. Thus, x^I cannot exceed:

$$1 - \frac{\gamma - \tau}{\gamma q^*(1 - \tau)}.$$

As a result, the equilibrium probabilities of media capture (x^I, x^E) solve the following constrained minimization problem:

$$\min_{(x^I, x^E) \in [0,1]^2} \left\{ x^I \pi_1 + q^*(1 - x^I)(ax^E + r(1 - x^E)) \right\} \tag{1.3}$$

$$s.t. \ x^E = 1 - \frac{\gamma - \tau}{(1 - x^I)(1 - \tau)\gamma q^*},$$

$$x^I \leq 1 - \frac{\gamma - \tau}{\gamma q^*(1 - \tau)}.$$

The effect of the prior on the probability of internal capture is non monotonous anymore. Without the positivity constraint on x^E, $x^I(\gamma)$ is a decreasing function: as in the main model, high priors favor internal capture. Thus, when γ is low, an increase in γ yields an increase in x^I, which needs to be compensated by a decrease in x^E to keep the credibility constraint satisfied. However, when the prior is large enough, $x^E = 0$ and it is not possible to decrease external capture further. Following an increase in the prior, the principal needs to reduce the probability of internal capture to remain credible.

Turning to the probability of observing external capture P^E, the effect of γ is now unambiguously negative. As long as x^E is positive, we have: $P^E = (1 - x^I)q^*x^E$. An increase in γ results in larger investigation efforts, which positively affects P^E. However, the positive effect on q^* is dominated by an increase in internal capture combined with a decrease in external capture to keep the credibility constraint satisfied.

These conclusions are summarized by the following results:

Proposition 5. *1. There exists a prior $\bar{\gamma}$ such that the probability of internal capture x^I decreases in the prior if and only if $\gamma > \bar{\gamma}$.*
2. The probability of external capture P^E decreases in the prior γ.

Proof. <u>1. Effect of γ on x^I.</u>
Starting from Equation 1.3, we use the expression of q^* as well as the positivity constraint to derive $\bar{x}^I(\gamma)$, the maximal value of x^I such that $x^E > 0$:

$$\bar{x}^I(\gamma) = \frac{\sqrt{(\gamma - \tau)(4a(\tau - 1)(\gamma - 1) - \tau + \gamma)} + \tau(2a\gamma - 1) - 2a\gamma + \gamma}{2a(\tau - 1)\gamma},$$

which is decreasing in γ.

Thus, as long as $x^I \in [0, \bar{x}^I(\gamma)]$, we can use the expression of q^* as well as the credibility constraint to rewrite the principal's objective as:

$$U_p(x^I) = \frac{a^2(\gamma((x^I - 4)x^I + 2) + x^I)}{2\gamma x^I - 2} - \frac{a(\gamma - \tau)}{(\tau - 1)\gamma} - \frac{r(\tau - \gamma)}{(\tau - 1)\gamma}.$$

80

$U_p(x^I)$ is a concave function on $[0, \bar{x}^I(\gamma)]$. Taking the F.O.C. and supposing that the positivity constraint does not bind, $U_p(x^I)$ is maximized for:

$$\hat{x}^I(\gamma) = \frac{1 - \sqrt{2}(1 - \gamma)}{\gamma},$$

which is increasing in γ.

As a result, the equilibrium probability of internal capture solves:

$$x^{I*}(\gamma) = \min\{\hat{x}^I(\gamma), \bar{x}^I(\gamma)\},$$

where $\partial \hat{x}^I(\gamma)/\partial \gamma > 0$, $\partial \bar{x}^I(\gamma)/\partial \gamma < 0$, $\lim_{\gamma \to \tau} \hat{x}^I(\gamma) < \lim_{\gamma \to \tau} \bar{x}^I(\gamma)$, $\lim_{\gamma \to 1} \hat{x}^I(\gamma) \geq \lim_{\gamma \to 1} \bar{x}^I(\gamma)$, which proves the claim.

2. Effect on P^E.

x^E is decreasing in γ and, as long as $x^E > 0$, P^E can be written:

$$P^E = \frac{\tau - \gamma}{\gamma - \tau\gamma} - \frac{(3\sqrt{2} - 4)\, a(\gamma - 1)}{2\gamma}.$$

Thus, we have:

$$\frac{\partial P^E}{\partial \gamma} = \frac{(4 - 3\sqrt{2})\, a + \frac{2\tau}{\tau - 1}}{2\gamma^2} < 0.$$

$\square$

B.4 Threats and violence

In the external capture stage, transfers are not necessarily positive: violence and intimidation of journalists are often used to influence news coverage, as we have seen in Figure 1.1. This appendix explores the consequences of such practices. We assume that the principal has to pay zt_i^E in order to make media outlet i incur a cost $-t_i^E$ if it publishes the signal. z is a random variable which depends on the institutional environment. It represents the relative cost of negative and positive transfers. It is publicly drawn after the investigation decision with $P[z = \bar{z}] = \alpha$, $P[z = \underline{z}] = 1 - \alpha$ and $0 \leq \underline{z} < 1 < \bar{z}$. We interpret α as a measure of protection of journalists.

The maximum number of media outlets externally captured depends on the realization of z. If $z = \bar{z}$ we still have at most $s_1 = \lfloor \frac{r}{a} \rfloor$ media outlets captured. However, when $z = \underline{z}$, external capture is cheaper and the principal can suppress up to $s_2 = \lfloor \frac{r}{\underline{z}a} \rfloor \geq s_1$ signals. We assume $n > s_2$.

If the principal uses threats, media outlets get a payoff of 0 when they receive a signal. The equilibrium efforts of investigation take the following form:

Lemma 3. *For $i > s_2 + 1$, $q_i^* = max\left\{ c'^{-1}(P[\theta = c|t^I]a/i), q_{s_2+1}^* \right\}$,*

$$\begin{cases} q_{s_2+1}^* \in \left[c'^{-1}\left(P[\theta = c|t^I] \frac{a\alpha}{s_2} \right), c'^{-1}\left(P[\theta = c|t^I] \frac{a}{s_2+1} \right) \right] if\ \alpha < \frac{s_2}{s_2+1}, \\ q_{s_2+1}^* = c'^{-1}\left(P[\theta = c|t^I] \frac{a}{s_2+1} \right) if\ \alpha \geq \frac{s_2}{s_2+1}, \end{cases}$$

$$\text{for } s_2 + 1 < i < s_1, \ q_i^* = \max\left\{c'^{-1}(P[\theta = c|t^I]a\alpha/i), q_{s_2+1}^*\right\},$$

$$\text{for } i \geq s_1, \ q_i^* = \max\left\{c'^{-1}(P[\theta = c|t^I]a\alpha), q_{s_2+1}^*\right\}.$$

Proof. Firstly, we consider $i = s_2 + 1$. If $q_{s_2+1} > c'^{-1}(P[\theta = c|t^I]\frac{a}{s_2+1})$, F.O.C. cannot be satisfied for $i = s_2 + 1$ and outlet $i = s_2 + 1$ would deviate for a smaller value.

In contrast with the main model, we can have $q_{s_2+1}^* < c'^{-1}(P[\theta = c|t^I]\frac{a}{s_2+1})$ due to the payoff discontinuity at q_{s_2}. In such a case, several media outlets pool on $q_{s_2+1}^*$ to escape the punishment.

However, we cannot have $q_{s_2+1} < c'^{-1}(P[\theta = c|t^I]\frac{a\alpha}{s_2}) < c'^{-1}(P[\theta = c|t^I]\frac{a}{s_2+1})$ in equilibrium: player $i = s_2$ would play $q_{s_2} > q_{s_2+1}$. $i = s_2 + 1$ would also deviate for a larger q_{s_2+1} (either until it equals $c'^{-1}(P[\theta = c|t^I]\frac{a}{s_2+1})$ and satisfies the F.O.C., or until it hits q_{s_2}, where there is the payoff discontinuity).

Thus, if $\alpha < \frac{s_2}{s_2+1}$,

$$q_{s_2+1}^* \in \left[c'^{-1}\left(P[\theta = c|t^I]\frac{a\alpha}{s_2}\right), c'^{-1}\left(P[\theta = c|t^I]\frac{a}{s_2+1}\right)\right]$$

can be part of an equilibrium.

Otherwise, if $\alpha \geq \frac{s_2}{s_2+1}$, $c'^{-1}(P[\theta = c|t^I]\frac{a\alpha}{s_2}) > c'^{-1}(P[\theta = c|t^I]\frac{a}{s_2+1})$ and there is a unique equilibrium where $q_{s_2+1}^* = c'^{-1}\left(P[\theta = c|t^I]\frac{a}{s_2+1}\right)$.

Given $q_{s_2+1}^*$, the equilibrium strategies of other media outlets are uniquely determined as follows:

- For $i > s_2 + 1$, $q_i^* = \max\left\{c'^{-1}(P[\theta = c|t^I]a/i), q_{s_2+1}^*\right\}$,
- For $s_2 + 1 < i < s_1$, $q_i^* = \max\left\{c'^{-1}(P[\theta = c|t^I]a\alpha/i), q_{s_2+1}^*\right\}$,
- For $i \geq s_1$, $q_i^* = \max\left\{c'^{-1}(P[\theta = c|t^I]a\alpha), q_{s_2+1}^*\right\}$.

$\square$

Compared to the main model, media outlets investigate less because of the punishment, which is in line with the "chilling effect" documented in the literature (Stanig 2015, Gratton 2015). The effect of competition is now ambiguous. Simultaneous publication implies to share profits, but it also helps to escape the punishment of the principal. In particular, suppose $\alpha < \frac{1}{s_2+1}$. There is an equilibrium where the first $s_2 + 1$ media outlets pool on the same investigation level to receive their signal simultaneously. If $v < s_2 + 1$, too many signals are received and the principal cannot deter publication. In such a case, investigation efforts are complement. If a media outlet deviates to a higher level of investigation, it could be the only one to receive a signal. However, the principal is likely to suppress this signal violently, which makes such a deviation not profitable. Due to a potential pooling at q_{s_2+1}, there can be multiple equilibria in which unilateral deviations are deterred by the risk of violent silencing.

Now, we turn to internal capture. Because of the possible complementarity of investigation efforts, several equilibria can exist for a given offer t^I. For example, suppose $\alpha < \frac{1}{s_2+1}$

and the principal offers $t_i^I = \pi_1^{\theta=c}$ $\forall i$. There is an equilibrium where all media outlets accept the offer. However, there is also an equilibrium where (at least) $s_2 + 1$ media outlets reject the offer. In this example, a joint deviation is possible because a coalition of $s_2 + 1$ media outlets can achieve more than a unique media outlet.

Thus, we need an equilibrium selection to continue the analysis. In response to an offer t^I, we suppose that media outlets play the equilibrium where the largest number of them reject the offer. This is equivalent to assuming that, if a coalition of media outlets can jointly improve their payoff by rejecting the offer of the principal, they would choose to remain free. Under this equilibrium selection, we have the following result:

Proposition 6. *Suppose $\alpha < \frac{1}{s_2+1}$. The principal internally captures k^* media outlets with some probability, where $n - s_2 \leq k^* \leq n$.*
$n - s_2$ media outlets receive $t_i^I = \pi_{s_2+1}^{\theta=c}$ and the others get $t_i^I = \pi_{n-k^*+1}^{\theta=c}$, with:

$$\pi_{s_2+1}^{\theta=c} = \max_q \{q\frac{a}{s_2 + 1} - c(q)\}$$

and

$$\pi_{n-k^*+1}^{\theta=c} = \max_q \{q\frac{a\alpha}{n - k^* + 1} - c(q)\}.$$

When protection of journalists is low, the logic of internal capture contrasts with Lemma 2. The principal does not only have to check individual deviations, he must also prevent a coordination from a critical number of media outlets. Thus, $n - s_2$ outlets receive relatively large payments to prevent the formation of such a coalition.

Once $n - s_2$ media outlets receive $t_i^I = \pi_{s_2+1}^{\theta=c}$, additional media outlets captured cannot form a large enough coalition to avoid the punishment of the principal and receive the expected payoff of a unilateral deviation. Unlike in the main setup, all investigation efforts of the first $s_2 + 1$ media outlets matter for the payoff of the principal. Depending on the parameters, he can buy any number of media outlets between $n - s_2$ and n. Finally, the media outlets that remain free cannot escape the retaliation of the principal. The expected value of a signal is small and their level of investigation is low.

C Additional Figures to Chapter 1
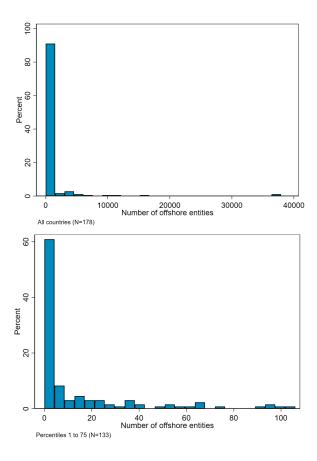
Figure C.1: Distribution of the number of offshore entities (country level)

Notes: An observation is a country. The second graph includes countries that belong to the first 75 percentiles of the distribution of offshore entities.
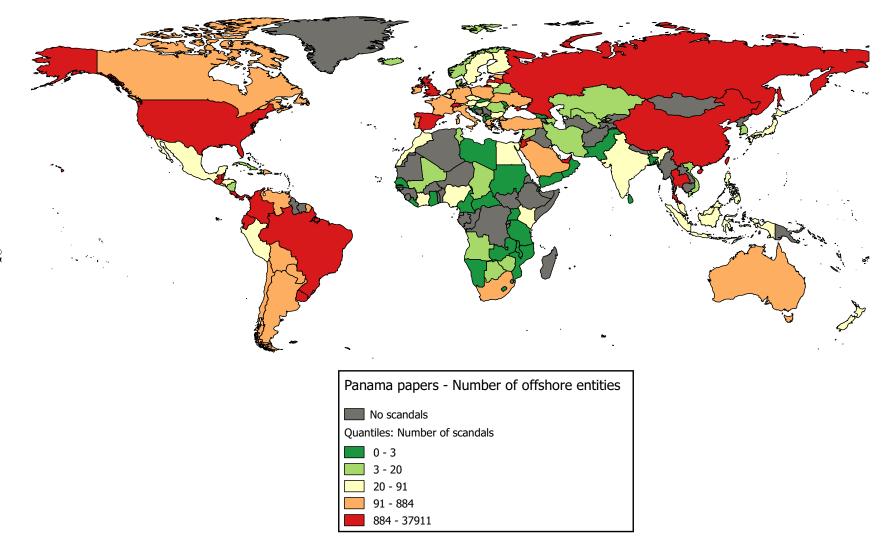
Figure C.2: Number of offshore entities revealed in the Panama Papers

a. Population weighted

b. GDP weighted

No recorded offshore entities
Weighted number of offshore entities (percentiles)
p0 - p19
p20 - p39
p40 - p59
p60 - p79
p80 - p99
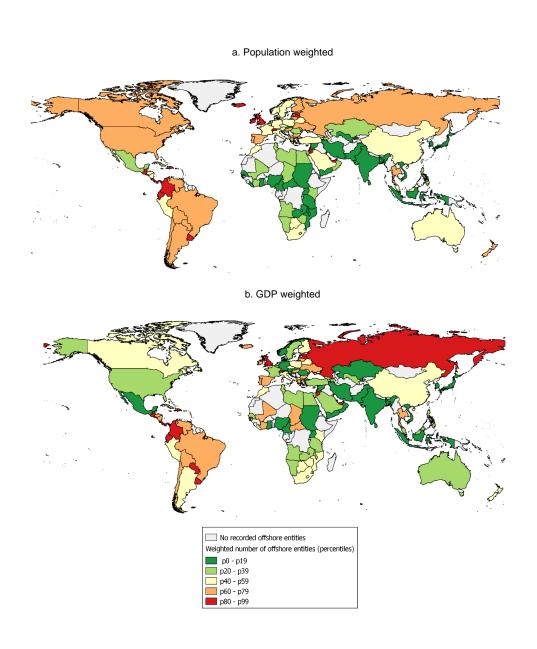
Figure C.3: Number of offshore entities revealed in the Panama Papers per thousand inhabitants or weighted by GDP
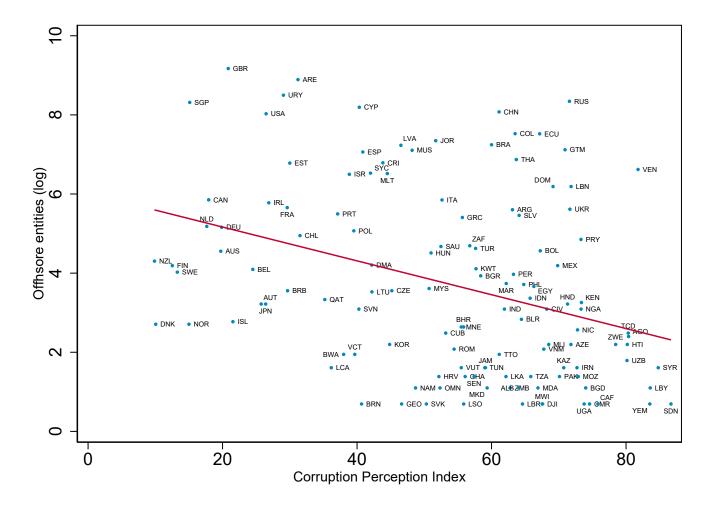
Figure C.4: Number of offshore entities (log) in the Panama Papers and perceived corruption

Notes: An observation is a country with at least one offshore entity mentioned in the Panama Papers (N=126). The Corruption Perception Index is averaged for the period 2012-2018. Coefficient of correlation: $-0,43^{***}$ $(R^2 = 0.12)$
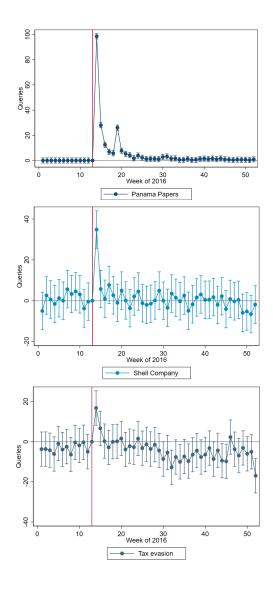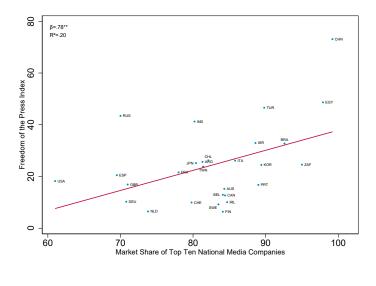
Figure C.5: Queries on Google Trends

Notes: An observation is a week. The red line marks the week that followed the publication of the files. Estimated equation: $word_{wc} = \alpha + \Gamma_w + \gamma_c + \epsilon_{wc}$ where $word_{wc}$ measures the number of Google queries of a given word in country c and week w. We collect the data in four languages: English, Spanish, French and Portuguese. The number of searches is expressed in relative terms based on the higher number of searches over the year. The figure plots the fixed effect estimates Γ. The omitted category is the week before the publication of the Panama Papers.
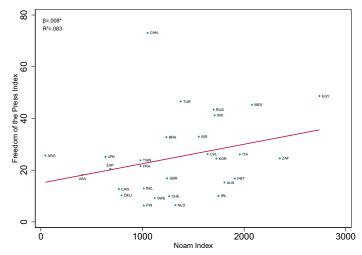
Figure C.6: Freedom of the Press Index and media concentration

Notes: Freedom of the Press Index is the score published by Reporters Without Borders in 2012. Market Share of Top Ten National Companies and Noam Index in 2011 – also referred to as the Media Ownership and Concentration Diversity Index (MOCDI) –, are from Noam (2016). The Noam index captures the degree of concentration of the market through the Herfindahl-Hirschman Index corrected for the number of players on the market.

$$Noam\ Index = \sum_{i}^{n} MarketShare_i^2 \times \frac{1}{\sqrt{n}}$$

D Additional Tables to Chapter 1

Table D.1: The Panama Papers and perceptions of corruption (tax havens excluded)

	Corruption Perc. Index			
	(1)	(2)	(3)	(4)
After × Exposure	1.225**		1.379*	
	(0.587)		(0.797)	
After × Exposure (log)		0.113		0.145
		(0.097)		(0.147)
Observations	1,113	1,113	1,031	1,031
Mean DepVar	58.60	58.60	57.10	57.10
Sd DepVar	19.30	19.30	18.98	18.98
Adj-R2	0.98	0.98	0.98	0.98
Year FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
After × Controls			✓	✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimation of Regression 1.1. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. The dependent variable is Transparency International's Corruption Perception Index. *Controls* include log GDP, growth rate over the period, population, and number of multinational companies based in the country. Luxembourg, Switzerland and Hong Kong are excluded.

Table D.2: The Panama Papers and perceptions of corruption (2012-2017)

	Corruption Perc. Index			
	(1)	(2)	(3)	(4)
After × Exposure	1.238**		1.671**	
	(0.586)		(0.771)	
After × Exposure (log)		0.086		0.167
		(0.086)		(0.122)
Observations	971	971	896	896
Mean DepVar	57.90	57.90	56.56	56.56
Sd DepVar	19.90	19.90	19.50	19.50
Adj-R2	0.99	0.99	0.99	0.99
Year FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
After × Controls			✓	✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimation of Regression 1.1. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. The dependent variable is Transparency International's Corruption Perception Index. *Controls* include log GDP, growth rate over the period, population and number of multinational companies based in the country. Data from 2018 are excluded.

Table D.3: The Panama Papers and perceptions of corruption, World Bank index

	World Governance Indicators Corruption			
	(1)	(2)	(3)	(4)
After × Exposure	0.900		1.150	
	(0.964)		(1.253)	
After × Exposure (log)		0.070		0.131
		(0.125)		(0.160)
Observations	1,133	1,133	1,045	1,045
Mean DepVar	55.00	55.00	52.97	52.97
Sd DepVar	29.73	29.73	29.07	29.07
Adj-R2	0.98	0.98	0.98	0.98
Year FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
After × Controls			✓	✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimation of Regression 1.1. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. The dependent variable is the Perception of Corruption from the World Governance Indicators. *Controls* include log GDP, growth rate over the period, population, and number of multinational companies based in the country. Luxembourg, Switzerland and Hong Kong are excluded.

Table D.4: Panama Papers and objective corruption

	Objective Corruption			
	(1)	(2)	(3)	(4)
After × Exposure	0.010		-0.013	
	(0.050)		(0.067)	
After × Exposure (log)		0.006		0.015
		(0.007)		(0.012)
Observations	987	987	917	917
Mean DepVar	0.27	0.27	0.27	0.27
Sd DepVar	0.30	0.30	0.30	0.30
Adj. R2	0.48	0.48	0.48	0.48
Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
After × Controls			✓	✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimation of Regression 1.1. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. The dependent variable is the measure of objective corruption constructed by Furceri et al. (2019). *Controls* include log GDP, growth rate over the period, population, and number of multinational companies based in the country.

Table D.5: Other forms of capture

	Advertising		Violence		Jail	
	(1)	(2)	(3)	(4)	(5)	(6)
After × Exposure	0.116**		0.772***		0.001	
	(0.046)		(0.273)		(0.070)	
After × Exposure (log)		0.009		-0.023		0.002
		(0.008)		(0.060)		(0.008)
Observations	875	875	875	875	875	875
Mean DepVar	0.57	0.57	5.72	5.72	0.25	0.25
Sd DepVar	0.39	0.39	2.52	2.52	0.35	0.35
Adj. R2	0.61	0.61	0.67	0.66	0.58	0.58
Country FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimation of Regression 1.1. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. Dependent variables - Source: Reporters Without Borders Experts Survey. *Advertising*: Does the government pressure advertisers to favour certain media? - *Physical violence*: Do journalists practise self-censorship for fear of the following consequences? Threats to physical safety of the journalist or his family and friends, to his workplace or his home - *Jail*: During the past 12 months, what penalties have been imposed on information providers? Prison sentence.

Table D.6: The Panama Papers and media capture with controls

	Internal capture				External capture					
	(1) Owners	(2) Owners	(3) Conflicts	(4) Conflicts	(5) In-kind	(6) In-kind	(7) Bribes	(8) Bribes	(9) Score	(10) Score
After × Exposure	1.355***		0.518		0.007		0.196		-0.535	
	(0.393)		(0.399)		(0.165)		(0.130)		(0.717)	
After × Exposure (log)		0.227***		0.046		0.049		0.065**		0.004
		(0.061)		(0.046)		(0.030)		(0.028)		(0.115)
Observations	875	875	875	875	863	863	875	875	907	907
Mean DepVar	5.36	5.36	6.84	6.84	4.53	4.53	4.10	4.10	32.67	32.67
Sd DepVar	2.22	2.22	2.19	2.19	0.88	0.88	1.16	1.16	15.35	15.35
Adj. R2	0.57	0.57	0.61	0.61	0.49	0.50	0.69	0.69	0.98	0.98
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
After × Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimation of Regression 1.1. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. Dependent variables: *Owners*: proportion of general-interest media owned by companies with interests in other economic sectors. *Conflicts*: likelihood of journalist self-censorship due to conflicts of interest with the owner. *In-kind*: non-monetary transfers to journalists to influence the coverage. *Bribes*: payments to journalists to influence the coverage. *Score*: Reporters Without Borders's aggregate score for the freedom of the press. *Controls* include log GDP, growth rate over the period, population, and the number of multinational companies based in the country.

Table D.7: The Panama Papers and media capture with controls (2012-2017)

	Internal capture				External capture					
	(1) Owners	(2) Owners	(3) Conflicts	(4) Conflicts	(5) In-kind	(6) In-kind	(7) Bribes	(8) Bribes	(9) Score	(10) Score
After × Exposure	1.160***		0.400		-0.038		0.055		-0.535	
	(0.362)		(0.423)		(0.174)		(0.123)		(0.717)	
After × Exposure (log)		0.218***		0.056		0.046		0.053*		0.004
		(0.059)		(0.047)		(0.030)		(0.028)		(0.115)
Observations	726	726	726	726	714	714	726	726	907	907
Mean DepVar	5.28	5.28	6.78	6.78	4.53	4.53	4.11	4.11	32.67	32.67
Sd DepVar	2.23	2.23	2.21	2.21	0.86	0.86	1.16	1.16	15.35	15.35
Adj. R2	0.60	0.60	0.60	0.60	0.48	0.48	0.69	0.70	0.98	0.98
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
After × Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimation of Regression 1.1. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. Dependent variables: *Owners*: proportion of general-interest media owned by companies with interests in other economic sectors. *Conflicts*: likelihood of journalist self-censorship due to conflicts of interest with the owner. *In-kind*: non-monetary transfers to journalists to influence the coverage. *Bribes*: payments to journalists to influence the coverage. *Score*: Reporters Without Borders's aggregate score for the freedom of the press. *Controls* include log GDP, growth rate over the period, population, and number of multinational companies based in the country. Data from 2018 are excluded.

Table D.8: External capture and competition on the TV market

	(1) In-kind	(2) Bribes	(3) In-kind	(4) Bribes
After × Exposure				
High Competition	-0.071	0.383*		
	(0.253)	(0.194)		
Low Competition	-0.409	-0.133		
	(0.278)	(0.251)		
After × Exposure (log)				
High Competition			0.084**	0.119***
			(0.036)	(0.033)
Low Competition			-0.011	-0.009
			(0.035)	(0.042)
Observations	567	577	567	577
Mean DepVar	4.52	3.99	4.52	3.99
Sd DepVar	0.89	1.13	0.89	1.13
Adj. R2	0.54	0.71	0.55	0.71
Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
After × Controls	✓	✓	✓	✓
pvalue: High=Low	.140	.006	.009	0.00

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimation of Regression 1.1, interactions with level of competition added. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. *High competition (resp. Low Competition)*: number of TV channels per 100,000 inhabitants above (resp. below) the median in the sample (m=4.9). *Controls* include log GDP, growth rate over the period, population, and number of multinational companies based in the country.

Table D.9: The Panama Papers and media capture (robustness panels)

	Internal capture				External capture			
	(1) Owners	(2) Owners	(3) Conflicts	(4) Conflicts	(5) In-kind	(6) In-kind	(7) Bribes	(8) Bribes
Panel A: No tax havens								
After × Exposure	0.866**		0.242		-0.163		-0.00477	
	(0.348)		(0.371)		(0.137)		(0.144)	
After × Exposure (log)		0.116**		0.0327		-0.00102		0.0249
		(0.0545)		(0.0487)		(0.0231)		(0.0228)
Observations	934	934	934	934	922	922	934	934
Mean DepVar	5.31	5.31	6.87	6.87	4.51	4.51	4.10	4.10
Sd DepVar	2.23	2.23	2.19	2.19	0.91	0.91	1.16	1.16
Adj. R2	0.53	0.52	0.59	0.59	0.50	0.50	0.64	0.64
Panel B: exclusion of USA, China, Russia								
After × Exposure	0.916***		0.243		-0.135		0.016	
	(0.350)		(0.371)		(0.138)		(0.144)	
After × Exposure (log)		0.148***		0.032		0.017		0.038
		(0.054)		(0.047)		(0.025)		(0.023)
Observations	932	932	932	932	920	920	932	932
Mean DepVar	5.28	5.28	6.84	6.84	4.49	4.49	4.06	4.06
Sd DepVar	2.26	2.26	2.20	2.20	0.91	0.91	1.17	1.17
Adj. R2	0.53	0.53	0.60	0.60	0.49	0.49	0.65	0.65
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimation of Regression 1.1. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. Dependent variables: *Owners*: proportion of general-interest media owned by companies with interests in other economic sectors. *Conflicts*: likelihood of journalist self-censorship due to conflicts of interest with the owner. *In-kind*: non-monetary transfers to journalists to influence the coverage. *Bribes*: payments to journalists to influence the coverage. Luxembourg, Switzerland and Hong Kong excluded from Panel A. USA, China and Russia excluded from Panel B.

Table D.10: The Panama Papers and media capture: ordered logit

	Internal capture				External capture			
	(1) Owners	(2) Conflicts	(3) In-kind	(4) Bribes	(5) Owners	(6) Conflicts	(7) In-kind	(8) Bribes
After × Exposure	0.948**	0.225	-0.636	-0.049				
	(0.431)	(0.484)	(0.431)	(0.376)				
After × Exposure (log)					0.185***	0.026	0.005	0.113*
					(0.070)	(0.060)	(0.059)	(0.060)
Observations	950	950	938	950	950	950	950	950
Mean DepVar	5.29	6.84	4.50	4.07	5.29	6.84	4.50	4.07
Sd DepVar	2.25	2.19	0.95	1.17	2.25	2.19	0.95	1.17
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Regression 1.1 estimated with an ordered logit. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. Dependent variables: *Owners*: proportion of general-interest media owned by companies with interests in other economic sectors. *Conflicts*: likelihood of journalist self-censorship due to conflicts of interest with the owner. *In-kind*: non-monetary transfers to journalists to influence the coverage. *Bribes*: payments to journalists to influence the coverage.

Table D.11: The Panama Papers and media capture (pre-trends)

	Internal capture				External capture			
	(1) Owners	(2) Owners	(3) Conflicts	(4) Conflicts	(5) In-kind	(6) In-kind	(7) Bribes	(8) Bribes
2012 × Exposure	-0.762		0.230		0.297		0.190	
	(0.468)		(0.445)		(0.225)		(0.264)	
2013 × Exposure	-0.499		0.927**		0.390*		0.351	
	(0.579)		(0.455)		(0.199)		(0.239)	
2016 × Exposure	1.178***		0.805*		0.283		0.377*	
	(0.442)		(0.477)		(0.238)		(0.219)	
2017 × Exposure	0.330		0.682		0.151		0.048	
	(0.484)		(0.550)		(0.256)		(0.227)	
2018 × Exposure	1.145*		1.048**		0.232		0.540**	
	(0.619)		(0.466)		(0.250)		(0.259)	
2012 × Exposure (log)		0.005		-0.090		0.064		0.011
		(0.078)		(0.062)		(0.046)		(0.048)
2013 × Exposure (log)		0.009		-0.077		0.035		-0.009
		(0.084)		(0.063)		(0.034)		(0.039)
2016 × Exposure (log)		0.228**		0.005		0.069*		0.056
		(0.092)		(0.069)		(0.039)		(0.045)
2017 × Exposure (log)		0.185**		0.005		0.066		0.032
		(0.088)		(0.081)		(0.047)		(0.045)
2018 × Exposure (log)		0.219**		-0.022		0.058		0.068
		(0.103)		(0.066)		(0.040)		(0.044)
Observations	864	864	864	864	852	852	864	864
Mean DepVar	5.40	5.40	6.87	6.87	4.54	4.54	4.12	4.12
Sd DepVar	2.21	2.21	2.18	2.18	0.88	0.88	1.14	1.14
Adj. R2	0.57	0.56	0.60	0.60	0.50	0.50	0.68	0.68
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Year × Controls	✓	✓	✓	✓	✓	✓	✓	✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimation of Regression 1.1. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. Dependent variables: *Owners*: proportion of general-interest media owned by companies with interests in other economic sectors. *Conflicts*: likelihood of journalist self-censorship due to conflicts of interest with the owner. *In-kind*: non-monetary transfers to journalists to influence the coverage. *Bribes*: payments to journalists to influence the coverage. *Controls* include log GDP, growth rate over the period, population, and number of multinational companies based in the country. *Exposure* and *Exposure (log)* interacted with all year dummies (excluding 2015).

Table D.12: 2SLS Estimation results

	OLS				1st stage	2SLS			
	(1) Owners	(2) Conflicts	(3) In-kind	(4) Bribes	(5) CPI	(6) Owners	(7) Conflicts	(8) In-kind	(9) Bribes
CPI	-0.00752	-0.0288	-0.0148	-0.0115		0.678	0.229	-0.0983	0.0345
	(-0.34)	(-1.46)	(-1.58)	(-1.17)		(1.53)	(0.68)	(-0.83)	(0.29)
After× Exposure					1.225*				
					(1.83)				
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	942	942	930	942	942	942	942	930	942
First-stage F-stat						10.47	10.47	11.91	10.47
Mean DepVar	5.27	6.84	4.50	4.06	58.60	5.27	6.84	4.50	4.06
Sd DepVar	2.24	2.19	0.91	1.17	19.30	2.24	2.19	0.91	1.17

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 2SLS estimation. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. Dependent variables: *Owners*: proportion of general-interest media owned by companies with interests in other economic sectors. *Conflicts*: likelihood of journalist self-censorship due to conflicts of interest with the owner. *In-kind*: non-monetary transfers to journalists to influence the coverage. *Bribes*: payments to journalists to influence the coverage. Columns 1-4 display the un-instrumented OLS estimates. Column 5 shows the first-stage results ($Corruption_{ct} = \Lambda_0 + \eta.After_t \times Exposure_c + \phi_t + \phi_c + \mu_{ct}$, the dependent variable is Transparency International's Corruption Perception Index, ϕ_t and ϕ_c year and country fixed effects). Columns 6 to 9 display the second-stage estimates ($Capture_{ct} = \Lambda_1 + \tau.\hat{Corruption}_{ct} + \phi_t + \phi_c + \varepsilon_{ct}$, $Capture_{ct}$ is one of the dependent variables and $\hat{Corruption}_{ct}$ is the value of the Corruption Perception Index predicted in the first stage).

Table D.13: External capture and quality of the media landscape

	(1) In-kind	(2) Bribes	(3) In-kind	(4) Bribes
After × Exposure				
Satisfactory	-0.062	0.211		
	(0.158)	(0.183)		
Intermediate	-0.050	-0.044		
	(0.152)	(0.151)		
Poor	-0.497**	-0.215		
	(0.243)	(0.237)		
After × Exposure (log)				
Satisfactory			0.019	0.053**
			(0.028)	(0.025)
Intermediate			0.027	0.018
			(0.029)	(0.029)
Poor			-0.040	0.016
			(0.052)	(0.040)
Observations	938	950	938	950
Mean DepVar	4.50	4.07	4.50	4.07
Sd DepVar	0.91	1.17	0.91	1.17
Adj. R2	0.50	0.66	0.49	0.65
Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
p-value: Satisfactory=Poor	.06	.08	.27	.39

Notes: $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$. OLS estimation of Regression 1.1, interactions with freedom of the press brackets added. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. Quality of the media landscape: *Satisfactory*, score between 0 and 25 - *Intermediate*, between 25.01 and 55 - *Poor*: above 55.01. Dependent variables: *In-kind*: non-monetary transfers to journalists to influence the coverage. *Bribes*: payments to journalists to influence the coverage.

Table D.14: The Panama Papers in countries in and out of the consortium

	Corruption Perc. Index			
	(1)	(2)	(3)	(4)
After × Exposure				
In the consortium	2.006**		2.732***	
	(0.835)		(0.997)	
Out of the consortium	1.700**		2.493***	
	(0.674)		(0.911)	
After × Exposure (log)				
In the consortium		0.083		0.154
		(0.097)		(0.139)
Out of the consortium		0.168		0.252
		(0.127)		(0.160)
Observations	1,203	1,203	1,101	1,101
Mean DepVar	57.15	57.15	56.00	56.00
Sd DepVar	19.68	19.68	19.23	19.23
Adj. R2	0.98	0.98	0.98	0.98
Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
After × Controls FE			✓	✓
pvalue: In=Out	.65	.53	.73	.51

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimation of Regression 1.1, interactions with participation in ICIJ added. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure*: categorical variable equal to 1 if country c is mentioned at least once in the Panama Papers leaks. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. The dependent variable is Transparency International's Corruption Perception Index. *Controls* include log GDP, growth rate over the period, population, and the number of multinational companies based in the country. *In (resp. out of) the consortium*: Treatment variable is interacted with a variable equal to 1 if at least one national outlet (resp. no outlet) was involved in the consortium of investigation.

Table D.15: The Panama Papers and advertising revenues

	Advertising			Circulation
	(1)	(2)	(3)	(4)
	All	TV	Newspapers	Sales
After × Exposure (log)	-0.001	-0.018	-0.006	0.004
	(0.018)	(0.017)	(0.017)	(0.012)
Observations	354	354	342	102
Mean DepVar	21.08	19.99	19.20	14.59
Sd DepVar	1.91	2.15	1.88	1.44
Adj-R2	0.98	0.99	0.98	0.99
Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimation of Regression 1.1. Standard errors in parentheses clustered at the country level. An observation is a country c in year t. *After*: variable equal to 1 after the shock, i.e. if $t \in \{2016, 2017, 2018\}$. *Exposure (log)*: logarithm of the number of entities from country c mentioned in the leaks. Dependent variables: *All media*, *TV* and *Newspapers*: logarithm of advertising revenues in USD (Columns 1, 2, and 3). *Sales*: logarithm of newspapers' revenues from sales (Column 4).

Chapter 2

Money and Ideology: Evidence from Candidate Manifestos

This chapter is joint work with Julia Cagé and Caroline Le Pennec.

Abstract

Do campaign contributions influence politicians? In this article, we study the impact of corporate donations on ideology and political discourse. We construct a novel dataset that combines the campaign manifestos issued by every candidate running for a seat in the French parliament with data on the amount and origin of their campaign contributions. We exploit an exogenous historical shock on corporate donations to estimate their causal impact on the content of campaign communication. Combining a difference-in-differences approach with computational text analysis, we show that receiving more donations from small and local corporate donors encourages candidates to advertise their local presence over national politics during the campaign. We also find evidence that donations lead candidates from extreme parties to moderate their rhetoric – including shifts in the policy topics they advertise. Our findings may reflect a "quid-pro-quo effect" between donors and politicians, but they may also result from an "electoral effect": receiving expressive contributions from corporate donors affects the content of campaign messages because it changes candidates' perception of which issues matter most to voters. According to our findings, campaign finance regulations may alter the information made available to voters through their impact on candidates' rhetoric.

1 Introduction

How do campaign contributions influence ideology and political discourse? In this paper we investigate the impact of corporate donations on the content of candidates' campaign communication. While a large body of the literature investigates the link between campaign spending and electoral success (Levitt, 1994; Ansolabehere et al., 2003; Kalla and Broockman, 2018) and studies the returns that firms can expect on their donations (Bombardini and Trebbi, 2011; Boas et al., 2014; Avis, 2020), the existing literature on special interest groups has overlooked the relationship between donations and political speech. However, campaign contributions are likely to affect not only votes and policies, but also the way candidates campaign, the ideas they promote during the electoral season and the information that is provided to voters before casting their vote.

The contribution of this paper is twofold. First, we exploit a national ban on corporate donations and combine a difference-in-differences approach with computational text analysis to investigate whether donations influence electoral discourse. We further test for heterogeneous effects depending on donor' type and the size of contribution, as well as candidates' political party. Second, we contrast our empirical findings with existing models of campaign finance and discuss the possible mechanisms that rationalize politicians' rhetorical response to corporate donations.

To carry out this analysis, we construct a novel dataset that combines data on the amount and source of the donations received by every candidate running for a seat in the French parliament, and the content of their communication with voters. French legislative elections are multi-party elections, in which candidates are elected using a two-round system in single-member constituencies. We use detailed information on each candidate's campaign revenue from Bekkouche and Cagé (2018), including the amount of corporate donations received by the 5,000 candidates running in 1993 – the last election held before the 1995 national ban that we exploit. We merge these data with the content of campaign manifestos, which all individual candidates can issue before the election and which are mailed to all registered voters by the State. Importantly, these two-page documents are distributed only a few days before the election, after most fundraising and campaign spending has finished. We use the manifestos issued before the 1993 election from Le Pennec (2020), and collect an entirely new corpus of manifestos issued before the 1997 election from paper archives. Our final dataset includes 10,285 single manifestos. We complement this candidate-level information with data on electoral results as well as the legislative activity of elected representatives: the written questions they issue to bring their constituents' concerns to the government's attention, and their interventions during legislative debates. Finally, we collect further information on

each contributing corporate donor in 1993, including their identity; their sector of activity; whether they contribute to multiple candidates' campaigns in the same local area or to candidates running in different parts of the country; and whether they give to a single political party or to candidates across the whole political spectrum.

We apply various methods of computational text analysis – borrowed from Gentzkow et al. (2019) and Bertrand et al. (2018a), among others – to the content of candidate manifestos to test four hypotheses. First, we investigate whether corporate donations push candidates to run a more local campaign – e.g. focused on their district – or a more national campaign – e.g. focused on party organizations and prominent politicians. Second, we test whether candidates use more polarized or more moderate language in response to donations. Third, we estimate the impact of corporate donations on the prevalence of some policy topics – such as the economy or homeland security – over others. Finally, we measure the overall originality of a candidate manifesto, relative to other manifestos from the same party, which we use as a proxy for the quality of a candidate's communication. Each of these measures represents a dimension of language that individual politicians – although they are mostly constrained by the national platform of the party they are affiliated with – have control over and may adjust in response to the campaign contributions they receive.

Because these campaign contributions are not randomly allocated across candidates, estimating their causal effect requires an identification strategy that handles their endogenous nature.[1] We first investigate the individual- and district-level determinants of corporate donations. In 1993, about a third of all candidates running for French legislative elections received at least one donation from a corporate entity. On average, men tend to receive more corporate donations than women, while incumbents and re-runners receive more donations than first-time candidates. We also find that candidates from mainstream parties receive more – and larger – donations, than "niche" and marginal parties' candidates. District-level characteristics, however, play a minor role in the allocation of donations.

Then, following Bekkouche and Cagé (2018), we exploit a natural experiment: the unanticipated 1995 ban on political contributions from legal entities,[2] which was enforced for the first time at the unexpected 1997 legislative elections.[3] We exploit the fact that can-

---

[1]The endogeneity of campaign spending is a key issue that has been variously tackled by the campaign finance literature, which mostly focuses on its impact on votes. To remove the unobserved candidates' heterogeneity, Levitt (1994) uses same-races repetition in U.S. Congressional elections, while Gerber (1998) instruments spending with variables affecting fundraising abilities, such as wealth levels. Later research has resorted to survey data (Jacobson, 2006) or field experiments (see Gerber, 2004, for a review). More recent papers have used national bans on corporate donations to estimate their impact on the allocation of public procurement contracts, e.g. Baltrunaite (2020).

[2]We refer to donations from legal entities as "corporate donations" in the remaining of the paper.

[3]Although the next legislative elections were not due until May 1998, President Jacques Chirac decided to dissolve the National Assembly in April 1997.

didates varied significantly in their reliance on such donations before the ban and estimate a difference-in-differences model to quantify the impact of receiving corporate donations in 1993. We control for the candidate characteristics that predict corporate donations as well as candidate and party-year fixed effects to capture the selection on unobservables. In particular, our within-candidate approach ensures that the estimated effect of donations does not stem from the selection of candidates with better (time-invariant) communication skills or better attributes to put forward in their manifesto. Further, we confirm our results with a nearest-neighbor matching estimator, thereby accounting for all observable differences between "treated" and "control" candidates.

We show that candidates who receive more corporate donations tend to emphasize their local presence in their manifestos. This effect is both statistically significant and economically meaningful: according to our estimates, a one-standard-deviation increase in the amount of corporate donations received by a candidate increases the relative prevalence of local references over national politics in her campaign manifesto by 16% of a standard deviation. The effect is mainly driven by an increase in the frequency of local references in the manifestos, but we also observe a decrease in the frequency of national references.

We further investigate heterogeneity across parties, and find that the impact of corporate donations on the prevalence of local references is positive for the five main running parties.[4] However, it is particularly strong for candidates running as independents, and for candidates affiliated with "niche" and extreme parties in our sample, i.e. the far-right and the Green party. Next, while we do not observe any significant effect on discourse polarization for mainstream parties, we show that donations also push more marginal candidates to moderate their rhetoric on the left-right scale. Turning to the policy topics covered in the manifestos, we find evidence that donations push candidates to favor economic issues – such as construction and amenities or business regulations – over social issues or foreign policy. Once again, these effects are stronger for independent candidates and candidates from niche parties. These heterogeneous effects suggest that small candidates with virtually no chance of winning the election are the ones who respond most strongly to corporate donations in their campaign communication. Lastly, we estimate a positive but smaller impact of donations on the originality of candidates' manifestos relative to the manifestos published by the other candidates from the same party.

We consider and discuss several mechanisms that could rationalize these empirical findings. First, corporate donations may result in more local references in the electoral discourse because of a "resource effect": an increased campaign revenue, no matter its sources, may

---

[4]The Communist party, the Socialist party, the Green party, the right-wing conservative party, and the far-right party.

enable politicians to run a better-quality campaign with a more targeted communication, better tailored to their local electorate. This interpretation is inconsistent with the observed heterogeneity across donors: the impact of donations on local prevalence is driven by small, local and non-partisan donors exclusively. Moreover, although these results are more correlational, we do not find any evidence that other sources of campaign revenue such as personal contributions or party contributions increase local references the same way corporate donations do. We conclude that money is not neutral: the identity of the donor matters.

Next, corporate donors may expect particularized benefits in exchange for their (service-induced) donations. This "quid-pro-quo effect" interpretation is consistent with the fact that local donors, with primarily local interests, push candidates to put forward their local presence in their manifesto – and advertise the influence they can exert at the local level to serve their donors' interests. Under this scenario in which candidates "pay back" their campaign donors after the election, we expect corporate donations to primarily alter the discourse and legislative activity of elected politicians in power. As we do not find any significant impact of corporate donations on either the quantity or the content of questions to the government and debate interventions among elected representatives, we cannot conclude that politicians favor their corporate donors once in power.[5] Hence, while we cannot rule out the existence of some quid-pro-quo relationship between corporate donors and politicians, such a mechanism is unlikely to drive *all* of the candidates' rhetorical response to donations.

Instead, our preferred interpretation of the results is that (expressive) donations have an "electoral effect" on candidates: receiving contributions from local corporate donors raises the salience of local and economic topics. As their perception of voters' most important issues changes, candidates adjust their persuasive campaign communication accordingly. This effect is particularly significant for marginal candidates and outsiders, whose prior beliefs regarding voters' preferences may not be as precise as mainstream candidates'. The existence of such a mechanism, which has been overlooked in the existing literature, suggests that connections with private corporations may alter politicians' behavior even in the absence of corruption and quid-pro-quo agreements.

Finally, we show that our results are robust to the use of a number of different specifications. In particular, we find that their magnitude and statistical significance do not vary when we introduce district×year fixed effects, additional time-varying district-level controls (including measures of the state of the economy at the district level such as the change in the unemployment rate) or when we control for differential time trends across candi-

---

[5]Note however that French MPs are supposed to represent the general interest and not the specific interest of their constituency, perhaps limiting the scope of any quid-pro-quo effect in the context of our study.

dates with different predicted levels of donations. Despite our attempts to control both for many observable factors and for time-invariant unobserved factors, our estimates might still be driven by time-varying unobserved covariates correlated with corporate donations. But although we cannot ultimately separate the effect of donations from other confounding factors, a causal interpretation of the results seems plausible. We also show that, although there is some substitution between corporate donations and other sources of funding (especially party contributions),[6] our estimated impact captures the effect of receiving more corporate donations specifically, not the effect of receiving fewer contributions from other sources.

Literature review This paper contributes to various strands of the literature. First, we contribute to the campaign finance literature that studies the impact of political donations. The focus of this literature has mainly been on the effect of contributions and campaign spending on electoral results (see among others Jacobson, 1978, 2006; Abramowitz, 1988; Green and Krasno, 1988; Gerber, 1998; Erikson and Palfrey, 1998; Cagé and Dewitte, 2018; Carvalho, 2020).[7] In the French context, Bekkouche and Cagé (2018) first used the 1995 ban on corporate donations to isolate the causal effect of political giving on vote shares.[8] Another strand of this literature studies the quid-pro-quo effects of campaign contributions. A common view is that firms are willing to influence political decisions by financing candidates' campaigns. Recent papers have found that donations facilitate access to elected officials (Kalla and Broockman, 2016) and influence the allocation of public procurement contracts (Titl and Geys, 2019; Baltrunaite, 2020; Gulzar et al., 2021). Others suggest that, on the contrary, campaign contributions do not buy significant political favors (Fowler et al., 2020). These studies – and, to a smaller extent, our own paper – relate to a wider literature on the value of lobbying and political connections (Fisman, 2001). In France, Bertrand et al. (2018b) have shown that politically connected CEOs tend to alter corporate employment decisions to help regional politicians in their reelection efforts, while Delatte et al. (2020) provide evidence that national representatives may exert influence in their local district to favor banks that helped them get reelected by bailing out local firms.[9] To the best of our knowledge, we are the first to investigate whether donations affect ideological cues and political rhetoric before the election – not only among elected politicians, but among *all* candidates. Interestingly, we find that marginal candidates who never get elected are

---

[6]Note that this substitution effect is partly driven by the existence of spending limits.

[7]See also Avis et al. (2017) who study the effects of campaign spending limits on political competition and incumbency advantage in Brazil.

[8]On campaign finance in the French context, see also Palda and Palda (1998); Foucault and François (2005); François et al. (2016).

[9]Beyond political giving, Bertrand et al. (2018a) find that charitable giving to non-profit organizations also buys corporate influence over policy-making.

actually the ones whose electoral discourse is most affected by contributions.

We also investigate the determinants of corporate donations, both at the candidate- and at the district-level. By doing so, we contribute to a very large empirical literature on the determinants of political donations (Gimpel et al., 2006; Gordon et al., 2007; Chamon and Kaplan, 2013; Bonica, 2014; Powell and Grimmer, 2016; Barber, 2016; McCarty et al., 2016; Rhodes et al., 2018; Fouirnaies and Hall, 2018; Teso, 2020) – including a narrower set of studies in the French context (François and Sauger, 2006; François and Phélippeau, 2015).[10] While the focus of the existing literature has mainly been on large donors, we investigate the heterogeneity of the effects depending on the size of corporate donors, and highlight the role played by small donors at the local level. We also question the role of quid-pro-quo motivations and conclude that electoral discourse is more likely affected by another, more novel, mechanism: expressive donations from donors who wish to support the candidates they like (Bouton et al., 2018) – which affect politicians' rhetoric through changes in their beliefs.

Finally, our paper contributes to the literature on campaign communication and political manifestos. While campaign messages have been shown to matter in voter decisions (Feltovich and Giovannoni, 2015; Kendall et al., 2015; Cruz et al., 2018), less is known about the determinants of their content, especially in the context of parliamentary systems in which the policy positions of individual candidates are tied to their national party platform.[11] Le Pennec (2020) uses candidate manifestos issued before the French legislative elections (which we also exploit in this paper) to show that high-quality politicians who are electorally weak in their district strategically advertise neutral non-policy issues instead of their party platform in order to win votes. This paper studies an overlooked determinant of campaign messages: the donations received by candidates. Our findings suggest that the regulation of campaign contributions – or lack thereof – may alter the information that is provided to voters before casting their vote, as it influences what candidates choose to advertise during the electoral season.

The rest of the paper is organized as follows. In Section 2, we provide background on campaign finance laws in France, introduce the new dataset we built for this study, and provide descriptive statistics. Section 3 discusses the determinants of corporate donations and presents our empirical strategy. We report the estimated impact of corporate donations on the content of candidate manifestos in Section 4, and discuss possible mechanisms in

---

[10]There is also a very large theoretical literature on the drivers of donations (see among others Grossman and Helpman, 1994).

[11]For an extensive review of the literature on policy positioning in party manifestos (not candidates), see Adams (2012). For a candidate-level analysis of positioning under different electoral systems, see Catalinac (2018).

Section 5. In Section 6, we perform a number of robustness checks and discuss the validity of our results. Finally, Section 7 concludes.

2 Empirical strategy

The French legislative elections are held every five years in all 577 constituencies – which are single-member districts – to elect members of the National Assembly (*"Assemblée nationale"*), the lower house of the French parliament. In this article, we focus on the 555 districts that are in metropolitan France, excluding the French overseas territories. In each of these districts, an average of 10 candidates compete for one seat.[12] In the 1990s' period we consider, about half of all running candidates were affiliated with one of five main party organizations: the Communist party, the Green party, the Socialist party, the right-wing conservative party (*Rassemblement pour la République*), and the far-right party (*Front National*). Candidates could also run for smaller issue-specific or regional parties, and about 30% of candidates chose to run as independents, without the endorsement of any party.

In this article, we construct a new dataset combining campaign donations and candidate manifestos for the 1993 and 1997 legislative elections. We complement these data with information on the activity of elected politicians during the subsequent legislatures. To do so, we construct and merge several datasets described in this section. We first present an overview of campaign finance regulation in France.

2.1 Campaign finance in France

French legislation on campaign and party financing has changed quite dramatically since the 1980s.[13] Financing rules are now stable and mainly focus on the following aspects of political finance: (i) public funding of campaigns (through the reimbursement of campaign costs), (ii) public funding of political parties, (iii) regulation of the donations to candidates and political parties, and (iv) campaign spending caps.

Laws passed in 1988 introduced direct public funding of parties as well as public reimbursement of candidates' campaign costs.[14] Candidates are allowed to make personal

---

[12]Formally, these elections follow a uninominal plurality rule with a runoff. If a candidate obtains the absolute majority in the first round, as well as a minimum of 25% of all the registered voters, then she is elected. If no candidate obtains the absolute majority in the first round, there is a second round where the two most-voted candidates and the candidates who obtained more than 12.5% of the registered voters can take part. The candidate who obtains a plurality of votes then wins.

[13]This section partly draws on Gunlicks (1993); Cagé (2018); Bekkouche and Cagé (2018).

[14]Laws no. 88-286 and no. 88-227 of 11 March 1988.

contributions to their own campaign,[15] to use contributions from their party, and to receive private donations – up to a cap. The 1990 law[16] created the *"Commission Nationale des Comptes de Campagne et des Financements Politiques"* (CNCCFP) which has been checking and approving the accounts of candidates' campaigns ever since. Every candidate running in legislative elections has to name a financial representative (*"mandataire financier"*) at most 12 months before the election date to handle her campaign finances – which, in practice, prevents politicians from fundraising long in advance – and to provide a detailed account of her spending and revenues to the CNCCFP within the six months following the election. However, candidates have no obligation to disclose their campaign revenue or the origin of any received contributions *before* the election.[17] Importantly, candidates may receive more donations than the amounts they actually spend, but if this is the case they have to transfer the remaining funds to a political party or to general interest foundations at the end of the campaign – the so-called *"dévolution"* – hence they cannot set funds aside for the next electoral campaign (Electoral law, articles L52-4 and L52-5). In other words, French candidates may only raise and spend money for a specific campaign, and under relatively tight scrutiny.

The 1995 ban on corporate donations   The law of 1995[18] marked an important change in party and election financing with the prohibiting of donations from legal entities: since then, only "natural" persons (i.e. individuals) have been allowed to make political donations – a reform that we exploit in the empirical analysis.[19]

Importantly, our own research through archived news articles from the national daily newspaper *Le Monde* indicates that this ban could not have been anticipated during the 1993 legislative elections campaign. The Socialist majority had passed a law regulating corporate donations only three years prior to 1993 and imposed new rules for greater transparency a few months before the election – requiring candidates to provide a detailed list of their corporate donors along with the amount they received.[20] Discussions of a ban on corporate donations were initiated by the newly elected right-wing government in Fall 1994,[21] in the wake of

---

[15]These personal contribution are reimbursed – up to 47.5% of the spending limit – by the State if the candidate obtains more than 5% of the votes in the first election round.

[16]Law no. 90-55 of 15 January 1990.

[17]Newspaper articles suggest that campaign contributors and corporate donors in particular were not commonly mentioned during the 1993 campaign that we study.

[18]Law no. 95-65 of 19 January 1995.

[19]The 1995 law introduced a few other changes – including a decrease in spending caps across districts – which are expected to affect all candidates in the same way, whether they received corporate donations in 1993 or not.

[20]Law of 29 January 1993.

[21]The 1988 laws were passed by Jacques Chirac's right-wing gouvernment in March. The legislative elections of June 1988, following François Mitterand's reelection, brought a win for the Socialist party (Michel Rocard's government). The right came back to power in the 1993 legislative elections.

multiple scandals involving campaign financing and conflicts of interest that emerged after the 1993 elections in spite of the new regulations. The first article in *Le Monde* mentioning the possibility of such a ban was published on November 30, 1994.[22]

2.2 Campaign revenues

In this article, we collect very detailed data on candidates' campaign revenues and expenditures, and in particular data on corporate donations that include the identity of the donor.

Total revenues For the 1993 and the 1997 legislative elections, we use data from Bekkouche and Cagé (2018) on each candidate's aggregate campaign expenditures as well as her campaign revenue and its main sources: corporate and individual donations, party contributions and personal contributions. Table 2.1 presents descriptive statistics for each election.

On average, candidates spent 20,397 euros in the 1993 legislative elections, compared to 14,607 euros in 1997. The drop in total spending (and revenues) can be explained by the 1995 ban on corporate donations but also by a decrease in the spending limit,[23] and by the fact that the 1997 election was called only two months ahead of time – following the dissolution of the National Assembly by the President – which limited candidates' ability to raise money for their campaign.[24] Sources of revenues and aggregate amounts are heterogeneous across parties. As shown in Figure 2.1, candidates from the two mainstream parties – the Socialist Party and the conservative right-wing party (named *"Rassemblement pour la République"* in 1993) – had higher revenues on average than candidates from other parties. In particular, they received on average more corporate donations in 1993, and the share of corporate donations in their total revenue is also higher (respectively 40% and 46%; see Appendix Figure D.2).

Corporate donations For the 1993 legislative election – the only election for which corporate donations were allowed *and* candidates had to disclose the amount they received from each corporate donor – we collect detailed information on the origin and amount of each of

---

[22]"M. Méhaignerie confirme le prochain dépôt d'un projet de loi sur le financement des partis politiques."

[23]From 1991 to 1995, candidates were allowed to spend up to 500,000 francs (121,000 euros) per election, and only 400,000 Francs for constituencies with fewer than 80,000 inhabitants. After 1995, the spending limit is composed of a flat rate and an additional amount depending on the size of the constituency. In 1995, candidates were allowed to spend up to 250,000 Francs (52,403 euros) per election plus 1 franc (0.15 euro) per constituent. The change from a flat function of the population size (below and above the 80,000 inhabitant threshold) to a linear relationship decreased the spending limit faced by all candidates.

[24]Jacques Chirac called the dissolution on April 21, 1997 – hoping to secure a stronger majority in Parliament – and the first round of the election to renew the National Assembly was held on May 25.

Table 2.1: Summary statistics: campaign spending and revenue
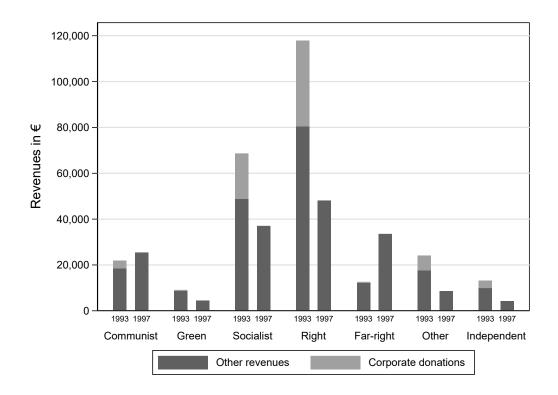
	Spending (cst euros)					
	Mean	Median	sd	Min	Max	N
Total spending per candidate						
1993	20,397	10,503	25,369	0	160,756	5,115
1997	14,607	2,257	18,646	0	72,122	5,977
Total revenues						
1993	22,923	10,583	33,326	0	784,482	5,134
1997	14,972	2,441	19,129	0	99,873	5,977
Share corporate donations						
1993	12.87	0.00	24.09	0	100	4,947
1997	0.00	0.00	0.00	0	0	5,026
Share individual donations						
1993	9.52	2.22	16.40	0	100	4,928
1997	13.81	0.05	25.47	0	100	5,001
Share personal contributions						
1993	35.78	15.97	39.49	0	100	4,926
1997	60.71	73.68	38.90	0	100	4,954
Share party contributions						
1993	37.81	20.56	40.06	0	100	4,924
1997	24.15	2.27	34.12	0	100	4,954

Notes: The table presents summary statistics on spending and revenues by candidates running in legislative elections. An observation is a candidate in 1993 or 1997. Revenues are measured in constant euros and shares in percentage points.

the donations made by corporations.[25] To do so, we digitize paper data from the official reports on election campaign costs and expenditures (*"Publication simplifiée des comptes de campagne"*) published by the CNCCFP. Appendix Figure D.1 provides an example of these data.

Table 2.2 displays summary statistics on the reported corporate donations. Approximately 33% of the candidates received at least one corporate donation (1,647 out of 5,141 candidates). The average number of donations received is equal to 2.85 when considering the whole sample of candidates (i.e. including those who received no corporate donation), and to 9 when we restrict the sample to candidates who benefited from corporate donations (see Appendix Table E.1 for summary statistics using this restricted sample). On average, candidates received 8,075 (constant) euros from corporate donors (0.12 euro per voter), and the mean revenue from corporate donations for candidates who received at least one donation

---

[25]We note here that donations were first allowed with the laws passed in March 1988, and candidates at the 1988 legislative elections that took place on June 5th and 12th were thus entitled to receive contributions both from individuals and corporations. However, they did not have to report their revenues or expenses to any centralized agency. See Appendix B for detailed information on the 1988 candidates' accounts.

Notes: This figure shows candidate mean revenues from corporate donations and other revenues by parties during the 1993 legislative electoral campaign. The sample includes all candidates endorsed by one of the five main parties (Communist, Socialist, Green, right-wing and far-right), as well as candidates affiliated to other smaller parties and independent candidates who are not affiliated with any party. Revenues are in constant euros. N= 11,372.

Figure 2.1: Revenues and corporate donations across parties

is 24,406 euros – which accounts for 37% of their total revenue (Appendix Table E.1).

The mean value of a donation is 2,061 euros; nonetheless, we observe significant variation in the distribution between large donations equal to the cap and very small contributions (part C of Table 2.2).[26] Heterogeneity can also be observed at the district level: in Appendix Table E.2, we provide summary statistics on the number of donations and the amount received across districts. Some districts are characterized by large flows of corporate donations (up to 109), and heterogeneity spans the whole territory as illustrated in Appendix Figure D.3.[27]

With regards to the corporate donors themselves, we first clean the firms' names and match them with administrative records to identify the donors that appear under different names or spellings and generate unique donor identifiers. Details about the procedure are presented in Appendix A.3. At the end of the procedure, we are left with 14,483 donations

---

[26]Corporate donations were capped by law at 50,000 francs – i.e. close to 10,000 euros (in constant euros 2017).

[27]In section 3.1, we further study the determinants of donations both at the candidate and the district levels.

made by 10,470 distinct donors. On average, donors gave a total amount of 2,857 euros to 1.38 different candidates (part B of Table 2.2). 84% of donors only gave to one candidate: we refer to these donors as *small donors* in the rest of the analysis.[28] Other corporations give to multiple candidates, possibly across many districts. These multiple donors contribute higher amounts on average: the mean donation of a small donor is equal to 1,469 euros while it is close to 3,000 euros among multiple donors. As an illustration, the public works company *COLAS* made 96 donations to candidates in 1993 for a total of 401,368 euros (140 times more than the average donor).[29] In Appendix Table E.3, we list the 20 corporations that made the highest number of donations.

We further classify these multiple donors along different criteria. First, we distinguish between single-district and multi-district donors, depending on whether all the donations of a corporate entity targeted candidates running in the same district. 81% of the multiple donors are multi-district donors (they represent 13% of the full sample of donors). Second, we distinguish between multiple donors who give only to right-wing or left-wing candidates, and those who give across parties: out of the 1,658 multiple donors, 38% only give to candidates from the same political orientation (10% to the left only and 28% to the right only).

Observing the number of donations or the type of the donor does not allow us to determine firms' motives in contributing to political campaigns. However, it is likely that multiple left-only or right-only donors are interested in pushing a specific partisan agenda. Conversely, donors who give across the political spectrum are likely not partisan.[30] Overall, these descriptive facts suggest that a large majority of corporate donors in our sample target individual candidates rather than partisan organizations. As additional evidence, we collect data on private donations to political parties (not individual candidates) from the paper-format reports of the CNCCFP and find that only 9% of corporate donors in our sample also contribute to a party. We further discuss donors' motivations in Sections 3 and 5, following existing models of political giving (Morton and Cameron, 1992) and distinguishing between service-induced donations (i.e. donors give to specific candidates to access power and receive particularized benefits in exchange) and expressive donations (i.e. donors give to specific candidates they want to support because they like their policy positions and/or

---

[28]These donors can take multiple forms, but a typical example is a small local business such as "Ets Bricchi Plomberie" (a plumbing firm).

[29]Further, this company also directly contributed to the funding of political parties across the political spectrum. For instance, in 1993, it gave 110,000 francs to the Communist party, 513,000 to the Socialist party, and respectively 255,000 and 610,000 to the two right-wing parties, as well as 900,000 francs to all the other parties taken together, totaling half a million euros (in today's constant terms).

[30]Noticeably, these *non-partisan* donors are likely to make more and larger donations: in Appendix Table E.4, we show that the number of donations and the mean donation of non-partisan donors are significantly larger. Among the top 20 largest donors, we do not record any left-only or right-only donors.

Table 2.2: Summary statistics: corporate donations in 1993

	Mean	sd	Min	Max	N
A. Candidates					
Corp. Donations > 0	0.33	0.47	0	1	5,141
# Corp. Donations	2.91	6.85	0	63	5,141
Corp. Donations in cst euros	8,075	20,738	0	330,208	5,141
Corp. Donation (euro/voter)	0.12	0	0	6	5,141
B. Donors					
# Corp. Donations	1.38	2.26	1	96	10,470
Total Donations	2,857	10,277	6	401,368	10,470
Small donor	0.84	0.37	0	1	10,470
Multiple donor	0.16	0.37	0	1	10,470
Single-district donor	0.03	0.16	0	1	10,470
Multi-district donor	0.13	0.34	0	1	10,470
Left-wing parties donor	0.10	0.30	0	1	1,658
Right-wing parties donor	0.28	0.45	0	1	1,658
Non-partisan donor	0.70	0.46	0	1	1,658
C. Donations					
Donation Value	2,061	2,561	6	9,842	14,483
Donation Value from small donors	1,469	2,175	6	9,842	8,811
Donation Value from multiple donors	2,981	2,831	10	9,842	5,672

Notes: The table presents summary statistics on corporate donations received by candidates in 1993. An observation is a candidate (part A), a donor (part B), or a donation (part C). Small donors are donors who made only one donation in 1993, multiple donors made more than one donation. Single-district donors (resp. multi-district) are donors who gave to candidates running in the same district (in more than one district). Left-wing parties (resp. right-wing parties) donors are multiple donors who made all the donations to candidates endorsed by left-wing (resp. right-wing) parties, non-partisan donors are multiple donors who gave to both left-wing and right-wing candidates. Donation value (part C) are in constant euros.

their attributes).

Last, we look at donors' sectors of activity. Given the format of the raw data, retrieving the activity is a challenging exercise. We provide details on the methodology we used to define the sector categories in Appendix A.3 and Appendix Table E.5. We identify the sector of half of the firms in our sample, and we label the others as "unknown sector".

2.3 Campaign manifestos and information on candidates

During the official campaign period, individual candidates have the right to issue one electoral manifesto (*"profession de foi"* or *"circulaire"*), which is distinct from their national party manifesto. These two-page documents are then mailed to all registered voters by the State a few days before the election, and surveys depict them as an important campaign tool. Importantly, manifestos are part of the official campaign spending that is fully reimbursed

by the state – provided that the candidate obtains at least 5% of the votes in the first round of the elections. Additional details can be found in Appendix A.1 and examples of candidate manifestos are provided in Appendix Figures D.5 to D.8.

Candidate manifestos issued before the 1993 legislative elections are from the Archelec project (CEVIPOF and SciencesPo Library) and Le Pennec (2020) (5,826 manifestos).[31] We gather 6,471 candidate manifestos issued before the 1997 election from the National Archives,[32] digitize the paper documents and apply optical character recognition to turn their content into machine-readable text.

We then use fuzzy string matching on candidates' names to merge the corpus with data on campaign donations, as well as candidate-level electoral outcomes from Bekkouche and Cagé (2018). The latter dataset provides information on the number of votes each candidate obtains in both election rounds, on her gender, on her political party, and on her other political mandates. Using their names, party labels and districts, we create a unique candidate identifier to follow candidates across elections.

2.4 Legislative activity

We collect information on the activity of the elected representatives from different sources. First, we collect the content of written questions to the government by scraping the National Assembly's website. These questions – which can be issued at any time even outside official legislative sessions – are directed to a single minister to express citizens' concerns on a specific topic.[33] We scrape the content of all the questions issued over the 9th, 10th and 11th legislatures, for a total of about 63,000 questions for the 1988-1993 period, 47,000 questions for 1993-1997, and 70,000 questions for 1997-2002.

Second, we scrape the content of representatives' interventions during public sessions. Unlike written questions, which are publicly available but remain relatively unknown to citizens, debate interventions have been broadcast on TV since the 1950s and may attract a large audience as well as media attention. The National Assembly's website provides the full record of these debates since the middle of the 10th legislature. Debates follow an opening question from a representative to a specific member of the government, to which other representatives can also respond. We parse the content of these conversations to isolate each single intervention and assign all interventions to their corresponding representative. We identify about 7,000 unique interventions over the 1993-1997 period and 20,000 interventions

---

[31]We also use, in some specification, the corpus of manifestos issued before the 1988, 1981, 1978, 1973, 1968 and 1967 elections, from the same sources.

[32]Classification numbers 19990140/32 to 19990140/36.

[33]See the general rules of the Parliament, Article 135.

over the 1997-2002 period.[34]

2.5 Text as data

We use computational text analysis to extract meaningful information from the textual content of candidate manifestos – and textual content once elected. We start with standard text pre-processing, which is described in Appendix A.2. We then construct four types of measures, which are presented and discussed below: the prevalence of local references over national ones, polarization on the left-right scale, the prevalence of different policy topics in discourse, and the originality of candidate communication relative to other candidates from the same party.

Local vs. national references First, we construct a simple measure of the attention each candidate gives to local aspects vs. national politics during the campaign. As candidates to a national parliamentary mandate, we expect these politicians to campaign on the national issues at stake. However, as they are running for election in a specific local district, we may also expect them to advertise their local presence in their campaign communication. To test whether corporate donations affect this balance between national and local campaigning, we count the number of times a manifesto mentions the department in which the candidate is running and the number of times it mentions a municipality (*"commune"*) located in that department, relative to the overall number of tokens in the manifesto. Although French *"députés"* hold a national mandate and are not responsible for local economic policy directly, the frequency of local references in one's manifesto is likely to reflect the candidate's stronger local ties – if she puts forward her participation to a municipal council for instance – and a better understanding of the local issues at stake. We also count the number of references to national politics, including the names of parties, party leaders and members of the government at the time of the election. We define the local index of a manifesto as the log ratio of its local frequency over its national frequency,[35] which measures the prevalence of local references over national ones in the document.

Appendix Figure D.9 shows the kernel density of this local index for each of the five main parties in our sample. On average, this index is negative, indicating that the frequency of local references (department and municipalities) tends to be lower than the frequency

---

[34]The content of debate interventions is not available for either the 9th legislature or the first years of the 10th legislature. Our sample of debate interventions starts on June 1, 1995, which explains why the number of interventions differs so much between 10th and 11th legislatures.

[35]More precisely, the local index is defined as $ln\left(\frac{1+Local}{1+National}\right)$, to take into account the multiple zeros in the frequency of national references.

of references to national politics in any manifesto. But there is some heterogeneity across parties, with a higher prevalence of local references for the Socialist, Green and right-wing parties, a slightly lower local index for the Communist party and a much lower one for far-right candidates.

Left-right partisan score and extremeness Next, we project the content of each document onto the left-right space of language. While most candidates are endorsed by a party and are tied to the policy platform decided at the national party level, they may decide to campaign on divisive partisan positions and issue a polarized manifesto, or to advertise consensus-based arguments and issue a more neutral manifesto instead.

We adopt a supervised approach to project all manifestos onto an ideological scale, leveraging the known party affiliation of candidates and the acknowledged ideological leaning of these parties from left to right. More precisely, we aggregate the content of manifestos issued by all candidates considered right-wing as well as the content issued by all candidates considered left-wing.[36] Then we give an ideological score to each word in the vocabulary, which reflects how likely a right-wing candidate is to use that word compared to a left-wing candidate. To do so, we follow the multinomial inverse regression approach proposed by Taddy (2013) and Taddy (2015), and we use a penalized estimator to estimate the model as recommended by Taddy (2017) and Gentzkow et al. (2019). All technical details can be found in Appendix A.2. Appendix Table E.6 shows examples of words with large negative loadings and words with large positive loadings in both 1993 and 1997. Left-wing words tend to refer to social policy ("poverty", "benefits") and capitalism ("dividend", "thatcher") while right-wing words refer to security issues ("terrorist", "murderer", "criminal"), immigration ("foreigner", "identity") and European integration ("europe", "independence").

The left-right partisan score of a manifesto is defined as the mean ideological score of the words it contains. Hence a document with a negative (positive) score is a document that uses primarily words used by politicians from the left (right) and rarely by politicians from the right (left), while a document with a score close to zero uses either polarized words from both ideological sides or uses neutral words that are used by politicians from both sides indifferently. Appendix Figure D.10 shows the kernel density of partisan scores (divided by their overall standard deviation), for each of the five main parties in our sample. We observe more extreme scores for Communist candidates (on the left) and candidates from the far-right (on the right) than for candidates from the more moderate Socialist party, Green party

---

[36]We determine whether a candidate's orientation is left- or right-wing using party affiliations and the political labels provided by the Ministry of the Interior, which assigns an ideological leaning to independent candidates as well (e.g. *Divers droite*). More precisely, we use the same classification as in Pons and Tricaud (2019).

and right-wing party. It suggests that candidates endorsed by more radical parties do indeed use more polarized language than others.

In addition to candidates' partisan leaning on the left-right scale, we define discourse extremeness as the absolute value of the partisan score – which measures the distance, either on the left or on the right, to a neutral manifesto.

Prevalence of different policy topics   We adopt a similar strategy to measure the prevalence of specific policy topics in campaign manifestos – and determine whether donations shift electoral discourse toward some topics more than others. The challenge is twofold, as this exercise requires first identifying such topics, and second measuring their relative importance in a given document. We do not know *ex ante* which candidates are more likely than others to talk about a certain topic, so we cannot use the manifestos themselves to build a supervised classifier as we do to scale manifestos from left to right. Instead, we use the set of all written questions to the government issued in the 9th, 10th and 11th legislatures as a training set, and the ministries targeted by these questions as topic labels. More precisely, we assign each ministry to one of four broad categories that are constant across legislatures: homeland security and administration, foreign policy, economy, and social issues – as well as a "non-classified" category. We perform a similar exercise with 17 narrower categories - homeland security, education, environment, retail, health, justice, economy, construction and amenities, public administration, employment, agriculture, defense and military, foreign policy, industry, culture, sport and entertainment, and European policy.

This method allows us to map political discourse, used by elected representatives once in office, onto topics that are most relevant to policy work as they represent the main government's activities. Using written questions is useful because the ministry they are addressed to is well-identified, and our sample contains more than 180,000 unique questions, allowing us to estimate the relationships between word usage and policy topics with high accuracy.

Once again, we follow the multinomial inverse regression approach proposed by Taddy (2013) and technical details are provided in Appendix A.2. Appendix Table E.7 shows examples of words with high loadings for each of the 17 narrow topics – translated into English.[37] Each manifesto is then represented as a set of probabilities, indicating the likelihood that the manifesto focuses primarily on a given topic over the others.

Appendix Table E.8 shows the mean and standard deviation of these predicted probabil-

---

[37]For instance, among the words most likely to be used to talk about homeland security, we find words referring to the organization of elections (for which this ministry is responsible), like "vote by proxy" and "electoral". We also find words referring to order and security – like "police", "firefighter", "violation" – as well as words referring to immigration issues, like "passport" and "algeria".

ities for each topic. The most prevalent topics are economy, education, employment, foreign policy and homeland security – with an average probability of dominating a candidate's discourse of 28 percentage points.[38] The prevalence of these topics can vary substantially across parties. As an illustration, Appendix Figure D.11 shows the distribution of homeland security prevalence in candidate manifestos for each of the five main parties in our sample. We see that candidates from the far right tend to use vocabulary associated with this policy topic much more often than candidates from any other party, especially in 1993. Conversely, candidates from the Green party focus less on homeland security than others.

Originality   Our last text-based outcome captures the originality of each manifesto relative to other manifestos issued by candidates from the same party. Anecdotal evidence from discussions with former candidates suggests that issuing a manifesto distinct from any party template can signal politician quality, beyond the policy platform they support. We note here that all three previous outcomes may already reflect deviations from the party line, because insisting on one's local presence, adopting a more polarized or moderate discourse, and choosing to emphasize some policy topics over others are all possible channels through which candidates can distinguish themselves from other candidates endorsed by the same party. To construct an overall measure of candidate originality that encompasses every possible channel, we follow Bertrand et al. (2018a) and use an unsupervised approach – Latent Semantic Indexing – which quantifies the similarity between two documents based on the words they contain (more details can be found in Appendix A.2). We define our originality index as the mean (negative) similarity between a candidate manifesto and every other manifesto from the same party. For this exercise, we restrict our sample to candidates affiliated with one of the five main party organizations in our sample and exclude independent candidates.

Appendix Figure D.12 shows the kernel density of candidate originality by party in 1993 and 1997 separately. While candidates from mainstream parties have relatively high levels of originality, suggesting that most of them tend to write their own personal manifesto, candidates from niche parties (the Green party particularly the far-right party) are more likely to use a template common to all candidates, resulting in very little originality among them – typically due to their limited experience as politicians, as compared to the candidates of mainstream parties. Interestingly, Communist candidates decrease in originality between 1993 and 1997 while Green candidates become more original over that period. This descriptive evolution is consistent with the decreasing influence of the Communist party on the left,

---

[38]This high number is partly explained by the fact that the Ministry of the Interior is responsible for organizing elections, and election logistics tend of course to be mentioned a lot in campaign manifestos.

and the growing success of a better-established Green party instead.

3   Impact of corporate donations on campaign communication

In this section, we study the determinants of corporate donations, provide evidence that such donations are not randomly allocated across candidates, and discuss how we address this endogeneity issue to estimate the causal impact of corporate donations on campaign communication.

3.1   What are the determinants of corporate donations?

First, we investigate the determinants of corporate donations. To do so, we estimate the following model:

$$\text{Corporate Donations}_{ipd} = W_i'\lambda + Z_d'\gamma + \eta_p + u_{ipd} \tag{2.1}$$

where the dependent variable of interest, Corporate Donations$_{ipd}$, is alternatively the number of corporate donations or the amount of corporate donations (in euros per voter) received by candidate i from party p in district d in 1993. W_i is a vector of individual-level covariates, including the candidate gender, and a number of indicator variables for whether (i) she ran in the previous election; (ii) she is the incumbent; (iii) she holds a mayoral mandate; (iv) she holds another political mandate (e.g. European MP or Senator). We also investigate whether campaign contributions depend on a text-based measure of candidate's ideological leaning, using her manifesto issued at the previous legislative elections (1988). Z_d' is a vector of district-level controls including the age composition, occupational structure and educational level of districts, as well as measures of their economic activity.[39]

We also include electoral controls such as the number of registered voters in the district and measures of past electoral competition: a dummy indicating whether the election was won in the first round and the electoral margin of the winner in 1988. Alternatively, we simply control for district fixed effects α_d.[40]   Finally, we include party fixed effects η_p – leaving candidates who are not affiliated with any of the five main parties as the omitted category – in all our specifications, so that our set of coefficients λ are within-party estimates

---

[39]Details about district-level covariates are presented in Appendix A.4. Appendix Table E.9 shows summary statistics.

[40]In this case, we estimate the following model: Corporate Donations$_{ipd} = W_i'\lambda + \alpha_d + \eta_p + u_{ipd}$.

(within-party and within-district estimates when we also control for α_d). We cluster standard errors at the district level.

Candidate-level determinants Figure 3.1 reports the candidate-level determinants of corporate donations in 1993 (estimation of equation (2.1)). In Figure 3.1a, the outcome of interest is the number of corporate donations. If we first consider party affiliations, we find that candidates from the Socialist party and the conservative right-wing party on average receive more corporate donations than the candidates who are running as independent and are not affiliated with any party (the omitted category), while candidates from the Green, Communist and far-right parties tend to receive less donations (all estimates are significant at the 1% level). In Figure 3.1b, we also show that the amounts (in euros per registered voter) received by candidates from mainstream parties are larger (respectively 0.11 and 0.38 euro per voter for the Socialist and right-wing parties). These findings are consistent with the results of Bekkouche et al. (2020) and the summary statistics presented in Figure 2.1.

Individual characteristics also significantly correlate with the number of donations. We find that, on average, men receive more corporate donations than women, while incumbents, mayors and re-runners and receive more donations than others.[41] Results are similar when we look at the amount candidates receive from corporate donors in Figure 3.1b. For instance, being a mayor increases candidate revenues from corporate donations by 0.3 euro per voter. It suggests that having a local foothold or better political connections is an important determinant of candidates' connections to firms at the fundraising stage of the campaign.

Previous findings from the literature highlight that candidate ideology is an important component in explaining fundraising (see e.g. Baron, 1994; Ensley, 2009). To test for the effect of ideological leaning – beyond party affiliation – we focus on the subset of candidates who run both in 1988 and 1993 and show the results in Appendix Figure D.13. We find a positive relationship between a candidate's left-right score measured at the previous election and the number of corporate donations received in 1993 – while still controlling for party fixed effects. It suggests that within the same party, candidates who use more right-wing language are likely to receive more corporate donations. However, the point estimate is small and only marginally significant.[42] Turning to the amount of donations received, Appendix Figure D.13b shows no relationship between the left-right score measured at the previous elections

---

[41]These results are consistent with Brollo and Troiano (2016) who document that, in Brazil, women incumbents benefit less from corporate campaign contributions when seeking reelection, as compared to male candidates.

[42]This specification also includes the candidate's vote share in 1988: the coefficient is small and insignificant, suggesting that once controlling for party affiliation and other measures of electoral success such as incumbency status, receiving more votes in the past does not predict more donations.

and the amount of donations received in 1993, suggesting that while party endorsement is a key determinant of corporate donations, within-party ideological leaning plays a minor role for donors.

Lastly, note that our candidate-level results are robust to a within-district analysis, in which we replace district level covariates Z'_d in equation 2.1 by district fixed effects α_d, leaves the coefficients on candidate characteristics almost unchanged (Appendix Figure D.14).
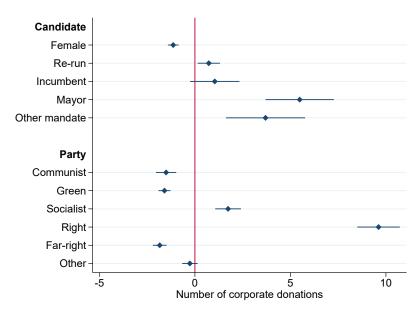
District-level determinants Appendix Figure D.15 suggests that district-level factors, such as demographic and occupational structure or economic activity, play a minimal role in determining the allocation of corporate donations across candidates, either at the extensive or at the intensive margin. We note, though, that the degree of electoral competition in the previous election may matter for donors' decision-making, as candidates running in districts that did not hold a runoff in the previous election (i.e. districts that are not very competitive because the front-runner was strong enough to win in the first round) tend to receive fewer than candidates running in more competitive districts.[43] In Appendix Figure D.16, we perform a similar analysis but consider the overall amount and number of corporate donations received at the district level by all the candidates. The results point toward the absence of major district-level drivers of corporate donations as well.

These descriptive findings provide some insights that we find worth noting. While our analysis of candidate-level determinants of donations (Figure 3.1) provides suggestive evidence that corporate donors favor politicians with better access to power (like incumbents and mayors),[44] the null results on district-level economic determinants provide a more nuanced picture, as corporate donors do not target politicians in places where their power would be most profitable to firms – e.g. in districts where municipalities have higher operating revenues and where more money can be spent on public contracts. Hence the primary motive of corporate donors is unlikely to be the pursuit of particularized benefits and economic favors from politicians, in exchange for their campaign contributions. Instead, the finding that more competitive districts attract more donations is consistent with donors contributing expressively to their preferred candidates' campaigns, especially when the outcome of the race is uncertain and when their support may make all the difference. We further discuss the relative importance of service-induced and expressive donations in Section 5.

---
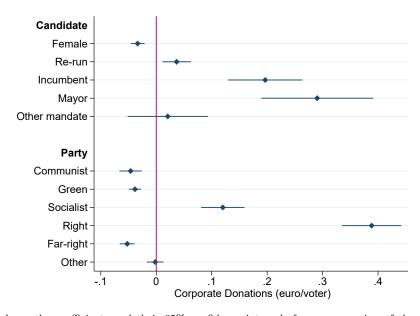
[43]The point estimate is relatively small (-0.03 euro per voter) but significant at the 1% level, and as large as the coefficients on the female and the re-run indicator variables.

[44]This pattern echoes findings in François and Sauger (2006), who argue that donors donate to seek influence based on the characteristics of the candidates they give to.

(a) Number of corporate donations

(b) Amount of corporate donations

Notes: This figure shows the coefficients and their 95% confidence intervals from a regression of the number of corporate donations (Figure 3.1a) or the amount of corporate donations in constant euros per voter (Figure 3.1b) received by each candidate on a set of party fixed effects (omitting independent candidates), candidate-level characteristics, and district-level characteristics (estimation of model 2.1). We use one observation per candidate in 1993. Standard errors are clustered at the district level.

Figure 3.1: Candidate-level determinants of corporate donations in 1993

3.2 Difference-in-differences approach

To control for the endogenous allocation of political donations, we exploit the ban on corporate donations passed between the 1993 and 1997 elections (see Section 2.2), and estimate the following model:

$$Y_{ipdt} = \alpha_i + \eta_{pt} + \beta \text{Corporate Donations}_{ipdt} + W'_{it}\lambda + u_{ipdt} \tag{2.2}$$

where α_i are candidate fixed effects, η_{pt} party×year fixed effects that control for party-specific time trends,[45] and Corporate Donations$_{ipdt}$ the amount of corporate donations per voter received by candidate i affiliated with party p and running in district d in year t (by definition, given the 1995 ban, Corporate Donations$_{ipd97} = 0$). Y_{ipdt}, the dependent variable of interest, is alternatively each of the text-based outcomes described in Section 2.5.

W'_{it} is a vector of time-varying candidate characteristics identified as strong determinants of corporate donations: indicator variables for having run for election before, being the incumbent, being a mayor and holding other electoral mandates. In addition to controlling for the selection of candidates on these observed characteristics, this specification controls for firms donating more often to individual politicians with specific time-invariant unobserved attributes, such as better communication skills. Our sample includes all candidates who ran both in 1993 and 1997.[46] We cluster standard errors at the district level.

Our identification assumption is that potential trends in campaign communication between 1993 and 1997 are uncorrelated with the allocation of corporate donations in 1993.[47] It is plausibly satisfied in our context. First, the 1995 ban on such donations was unexpected, preventing candidates with specific attributes and communication skills from raising large amounts of corporate donations in anticipation of their future loss in campaign revenue. Note furthermore that, as explained in Section 2.1, the French campaign finance rules prevent candidates from stockpiling funds for future campaigns once the election is over.

Second, the inclusion of time-varying controls in equation 2.2 captures any potential differences in communication trends caused by changes in the observed determinants of do-

---

[45]We include a separate fixed effect for each party organization, including smaller ones that are not among the five main party organizations, and a common fixed effect for independent candidates.

[46]Note that this sample of repeating candidates is a selected sample, an issue we tackle in Appendix C. As shown in Appendix Table E.10, 46% of these re-runners received at least one corporate donation in 1993, resulting in an average number of 5 donations and an average amount of 0.22 euro per voter in this sample.

[47]A classical approach to test for the validity of this assumption is to show that trends in outcomes were parallel before the policy change. Unfortunately, in our context, this exercise is compromised by the series of campaign financing reforms that preceded the ban on corporate donations in the late 1980's, including the introduction of (legal) corporate donations in 1988 (see Appendix B), and the stricter regulation of such donations in 1990 (see Section 2.1). Further, we would need the same candidates to run repeatedly in the same constituency in the 1980s and 1990s, which is rarely the case. We provide more details in Section 6.1.

nations. In Section 6, we further discuss the validity of our empirical strategy and show that our main results are robust to a wide range of alternative specifications – like including more time-varying controls, or controlling for differential trends across candidates with different *predicted* levels of corporate donations.

4   Impact of corporate donations on campaign communication

4.1   Main results

Anecdotal evidence   To illustrate the impact corporate donations may have on campaign communication, we first provide and compare two concrete examples. Appendix Figures D.5 and D.6 show the campaign manifestos issued by two different Green candidates in 1993. Monique Mascret (Appendix Figure D.5) received more than 10,000 euros in corporate donations and issued a rather personal manifesto in which she highlights her family, her occupation and her local anchoring – emphasizing the fact that she has lived in the district for 18 years. She advertises the key policy positions of the Green party regarding waste management and pollution, with very concrete proposals such as subsidizing farmers who reforest their land. Interestingly, she also advocates for pro-business economic policies, including the reduction of corporate taxes and the support of construction projects to boost employment. Conversely, Sophie Bouchard (Appendix Figure D.6) did not receive any corporate donation in 1993 and issued a more generic manifesto that highlights the core values of the Green platform (productivism, pollution, redistribution) without any concrete application, and provides very little information about the candidate herself or her background.

Next, Appendix Figures D.7 and D.8 show the campaign manifestos issued by two different candidates endorsed by the far-right party. Jacques Peyrat (Appendix Figure D.7) received close to 16,000 euros in corporate donations in 1993 and issued a manifesto that mixes proposals from the national platform of the party (immigration, tax reduction and conservative moral values) and a local corruption scandal involving the misuse of public funds by a previous mayor. Conversely, Ferdinand Ginoux (Appendix Figure D.8) did not receive any corporate donation and used a manifesto template that was common to almost all far-right candidates that year, with very little personalization. This template describes the national party platform and its most controversial policy proposals (such as reenacting the death penalty) and attacks all the other parties for their alleged political failures.

These two examples suggest that candidates endorsed by the same party may choose to highlight different types of arguments in their campaign communication, in particular

when comparing candidates who received corporate donations to those who did not. We now use the difference-in-differences approach described in Section 3.2 to characterize the impact of donations on campaign communication more systematically, while controlling for the endogeneity of donations across politicians. Table 4.1 presents the results.

Local anchoring   We first estimate equation (2.2) using the local index of each candidate manifesto, which measures the prevalence of local references over national ones, as the dependent variable.[48] As shown in column 1, a one-standard-deviation increase in corporate donations increases the local index by 16% of a standard deviation, an estimate that is significant at the 1% level.

Columns 2 and 3 show that this effect is mostly driven by a significant increase in the frequency of local references in manifestos (a 25-percentage-point increase that corresponds to about 25% of the mean local frequency after the ban), and a smaller decrease in the frequency of national references (a 13-percentage-point decrease that corresponds to about 5% of the mean national frequency after the ban).[49] Overall, these results suggest that receiving more donations encourages candidates to advertise their local presence in their campaign communication – either by referring to very specific local issues or by mentioning any other electoral mandate that they may hold in the department – while also reducing the number of references to national politics.[50]

Partisan leaning   Second, we test for the impact of corporate donations on partisan leaning in campaign discourse. Columns 4 and 5 of Table 4.1 show no significant impact of receiving donations on either the left-right score of candidate manifestos, or their extremeness – defined as the absolute value of the left-right score. It suggests that corporate donations do not systematically shift electoral discourse toward one ideological side or the other, and does not increase or decrease polarization either.[51] However, without necessarily adopting

---

[48]In this specification and all that follow, the local index is divided by its (yearly) standard deviation. We use the standardized aggregate amount of corporate donations as our main explanatory variable and restrict the sample to observations for which both the aggregate amount and the detailed breakdown of corporate donations are available. We show in Appendix Table E.11 that our results are robust to including all candidates for whom the aggregate amount of corporate donations is known, or to observations for which the reported aggregate amount of corporate donations is exactly equal to the sum of single donations.

[49]Note that these opposite effects are not mechanical: a candidate could increase both the number of local references *and* the number of national references in her manifesto – at the expense of any word that is neither a local keyword nor a national keyword.

[50]Consistent with these results, we find an overall decrease, among *all* running candidates, in the local index (13% of a standard deviation) and the frequency of local references (23 percentage points) after the ban. There is, however, no significant change in the share of national references between 1993 and 1997.

[51]We note that the positive impact on the frequency of local references should somewhat mechanically decrease polarization, as references to departments and municipalities tend to be neutral words by construction. This may explain the negative coefficient in column 5 of Table 4.1, but the fact that it is small and

Table 4.1: Impact of corporate donations on campaign communication

	Local index	Local references	National references	Left-right score	Extremeness	Originality index
	(1)	(2)	(3)	(4)	(5)	(6)
Corporate donations	0.162***	0.253***	-0.130***	-0.001	-0.004	0.033**
	(0.030)	(0.054)	(0.050)	(0.005)	(0.004)	(0.014)
Observations	2602	2602	2602	2602	2602	2070
Mean outcome after ban	-0.658	1.199	2.658	-0.021	0.740	-2.152
R2-Within	0.030	0.024	0.008	0.004	0.008	0.007

Notes: Standard errors are clustered by district and shown in parentheses (***, **, * indicate significance at 1, 5, and 10 percent, respectively). We use one observation per candidate per year. The sample includes all candidates who run both in 1993 and 1997, whose manifesto is available, and for whom both the aggregate amount and the detailed breakdown of corporate donations are known. In column 6, the sample is further restricted to candidates affiliated with one of the five main political parties (Communist, Socialist, Green, right-wing, or far-right party). We control for candidate fixed effects and party×year fixed effects, as well as time-varying individual controls: indicator variables for having run in the past, for being the incumbent, and for holding other electoral mandates. The amount of corporate donations per voter is divided by its standard deviation in 1993. The local index measures the prevalence of local references over national ones and is divided by its standard deviation for interpretability (column 1). The normalized frequencies of local and national references in candidate manifestos (columns 2 and 3) are measured in percentage points. The left-right score (column 4) measures the prevalence of right-wing (positive score) vs. left-wing (negative score) language in candidate manifestos, and extremeness is defined as the absolute value of the left-right score (column 5). The originality index is the mean (negative) similarity between each manifesto and all other manifestos issued by candidates from the same party in the same year, divided by its standard deviation (column 6).

133

more divisive or consensus-based language, corporate donations may affect the policy topics covered by the candidates.

Policy topics   In Table 4.2, we estimate the impact of corporate donations on the policy topics candidates advertise in their manifesto. As previously described, we aggregate the policy topics in four main categories: homeland and administration, foreign policy, economy, and social.

We show that a one-standard-deviation increase in corporate donations raises the probability of focusing on economic issues by 1.6 percentage points, an estimate that is significant at the 1% level and corresponds to a 5% increase relative to the mean prevalence of economic issues after the ban (column 3). Conversely, column 4 shows a negative effect on social issues (-1 percentage point), significant at the 10% level. We also obtain a negative impact (significant at the 5% level) on foreign policy (column 2). Finally, there is no effect on the probability of focusing on homeland security and administration (column 1).

Overall, these results suggest that corporate donations encourage candidates to devote more space in their campaign communication to economic issues, at the expense of foreign policy and social issues.

In Appendix Figure D.17, we report the point estimates and their 95% confidence intervals for the effect of corporate donations on the prevalence of 17 narrower topics. While few coefficients are statistically significant, the results suggest that the positive impact on economic issues is mostly driven by an increase in the prevalence of construction and amenities – with an estimated effect corresponding to 15% of a standard deviation, significant at the 1% level – and, to a smaller extent, by an increase in the prevalence of agricultural issues, retail, the economy (which refers mostly to market regulations) and the environment. Interestingly, the effects are not driven by *any* economic question, but by local issues like construction and amenities. On the contrary, economic topics that are more likely to refer to national policies – such as the industry – are negatively impacted by corporate donations (although not significantly).

Originality   Finally, column 6 of Table 4.1 shows a significant and positive impact of corporate donations on the originality index (3.3% of a standard deviation) – defined for candidates of the five main parties in our sample. This last result suggests that corporate donations push candidates to issue a manifesto that is distinct from the manifestos published by other candidates from the same party. It is consistent with the large impact on local

---

insignificant suggests that a deeper change in content and vocabulary would be needed to detect an overall moderation effect.

Table 4.2: Impact of corporate donations on broad policy topics

	Homeland and administration	Foreign policy	Economy	Social
	(1)	(2)	(3)	(4)
Corporate donations	0.508	-0.280**	1.572***	-1.140**
	(0.515)	(0.117)	(0.531)	(0.554)
Observations	2602	2602	2602	2602
Mean outcome after ban	15.189	3.011	28.788	34.827
R2-Within	0.005	0.004	0.014	0.007

Notes: The outcome is the predicted probability, for each policy topic, that a candidate manifesto focuses primarily on that topic – based on the words it contains. It is measured in percentage points. Other notes as in Table 4.1.

references – which distinguishes candidates from politicians running for the same national party but in other departments – and the impact on the prevalence of different policy topics. Note, however, that there is not a one-to-one mapping between these different outcomes: a candidate may insist on her local anchoring in one part of her manifesto but use a template identical to all other candidates from her party in other parts, yielding both a high local index and a relatively low originality index – and a possibly high extremeness score if that party template advertises polarized policy topics. Hence the relatively small impact on candidate originality compared to the effect on the local index and the frequency of local references suggests that advertising one's local presence is the main response to receiving corporate donations – not just an attempt, among others, to appear more original.

4.2 Heterogeneity of the effects

Depending on the political parties We now investigate whether the impact of corporate donations differs across parties. Table 4.3 shows the results from interacting Corporate Donations$_{ipdt}$ with seven indicator variables, indicating which party endorses candidate i – the Communist party, the Green party, the Socialist party, the conservative right-wing party, the far-right party or any other smaller party – or if the candidate is running as an independent. As shown in column 1, the impact on the prevalence of local references over national ones is positive for the five main parties.

The estimated impact of donations on the local index is particularly high for parties that are newer on the political scene: a one-standard-deviation increase in corporate donations per voter is estimated to raise the local index by 3.5 standard deviations among Green candidates (a huge effect that is significant at the 1% level) and by 49% of a standard deviation among far-right candidates (although it is not significant). As shown in columns 2 and 3, this effect is driven by a very large decrease in the frequency of national references for

Green candidates (-7 percentage points), though the impact of donations on the frequency of local references remains strong (0.8 percentage point) and significant at the 1% level. This effect on local references specifically is even larger for far-right candidates (1.3 percentage points, also significant at the 1% level).

Table 4.3: Heterogeneity by party

	Local index	Local references	National references	Left-right score	Extremeness	Originality index
	(1)	(2)	(3)	(4)	(5)	(6)
Communist*Corp.Don.	0.223***	0.337**	-0.233***	-0.016	0.013	0.031
	(0.068)	(0.166)	(0.066)	(0.016)	(0.016)	(0.047)
Green*Corp.Don.	3.478***	0.827***	-7.174***	0.628***	-0.620***	0.013
	(0.390)	(0.303)	(0.897)	(0.112)	(0.126)	(0.229)
Socialist*Corp.Don.	0.167***	0.227**	-0.174**	0.004	0.000	0.010
	(0.050)	(0.092)	(0.077)	(0.009)	(0.008)	(0.026)
Right*Corp.Don.	0.134***	0.215***	-0.089	-0.002	-0.005	0.044***
	(0.039)	(0.064)	(0.077)	(0.007)	(0.006)	(0.015)
Far-right*Corp.Don.	0.487	1.307***	0.141	-0.514	-0.536	0.304
	(0.647)	(0.309)	(1.506)	(0.490)	(0.489)	(0.993)
Other*Corp.Don.	-0.759	-1.002	0.420*	0.028	-0.121	
	(1.079)	(2.628)	(0.221)	(0.048)	(0.098)	
Independent*Corp.Don.	0.344**	0.657***	-0.171	0.007	-0.038**	
	(0.139)	(0.176)	(0.286)	(0.022)	(0.018)	
Observations	2602	2602	2602	2602	2602	2070
Mean outcome	-0.627	1.287	2.740	-0.027	0.804	-2.203
R2-Within	0.035	0.028	0.011	0.007	0.013	0.008

Notes: The amount of corporate donations per voter (divided by its standard deviation in 1993) is interacted with indicator variables indicating if the candidate is endorsed by any of the five main parties, by another smaller party or if the candidate is running as an independent. Other notes as in Table 4.1.

Interestingly, donations also affect the partisan leaning of discourse among candidates from these two parties: as shown in column 4, while the effect on the left-right score remains small and mostly insignificant for the three historically dominant parties (right-wing, Socialist and Communist), Green candidates move to the right of the language scale and far-right candidates move to the left (although the estimate is only significant for the Green candidates). As a consequence, receiving corporate donations results in less extreme campaign messages among both Green candidates and, to a smaller extent, far-right candidates (column 5).

Overall, this heterogeneity suggests that the rhetoric of marginal parties – which are not

yet part of governing coalitions – responds more strongly to campaign contributions, as compared to well-established parties.[52] This is also true of independent candidates – who are not endorsed by any party, with a positive effect of 0.7 percentage point on the frequency of local references (significant at the 1% level) and a small but significant moderation effect (3.8% of a standard deviation). Appendix Table E.12 corroborates this pattern: the positive effect of donations on the prevalence of economic issues and their negative effect on social issues are also higher for niche and independent candidates than for mainstream ones (although not all estimates are significant).

Finally, column 6 shows that there is no systematic relationship between a strong impact on local references (or a strong moderation effect) and a strong increase in manifestos' originality. Once more, these findings suggest that the observed effects of corporate donations cannot be summarized as a single originality effect.

Depending on the candidates' characteristics   Finally, we explore heterogeneity across different types of candidates. Columns 1 through 3 of Appendix Table E.13, in which we estimate a version of equation 2.2 where Corporate Donations$_{ipdt}$ is interacted with each of the individual characteristics included in model 2.1 – measured in 1993 – shows that the impact of corporate donations on the prevalence of local references over national ones is stronger among candidates who were incumbents, mayors or held other electoral mandates at the time they received donations (although none of these coefficients is significant at any conventional level). The estimated effect on the frequency of local references is particularly strong for candidates with other mandates (column 2). This result is intuitive, as adapting one's campaign communication toward a more local message is likely easier for experienced politicians who are already in power, who have better local connections as well as a record of local achievements that they can put forward in their manifesto.

5   Mechanisms

To summarize our findings, we have provided evidence that corporate donations affect politicians' campaign communication: they push candidates to advertise their local presence over national politics, especially candidates from niche parties with extreme or issue-specific platforms. These candidates are also found to moderate their discourse in response to do-

---

[52]Seen from the 21th-century perspective, one might be surprised by the fact that we here consider the Communist party to be a well-established party, while the Green party and the far-right party (Front National) are presented as outsiders. However, in 1990s' France, the Communist party was still an important political party, and Ministers from the Communist party were governing during François Mitterrand's presidency.

nations. In addition, the policy topics candidates choose to emphasize in their manifestos are likely to be affected by the amount of corporate donations they receive. We now discuss several interpretations and possible mechanisms for these empirical results – especially the strong effect of corporate donations on the prevalence of local references.

We consider several hypotheses that could rationalize our results: (i) corporate donations increase revenue and allow candidates to run a better campaign; (ii) politicians "pay back" their donors and adapt their political agenda to serve their interests, which is reflected in their campaign communication; (iii) fundraising and building connections with corporate donors affects candidates' beliefs regarding on voters' preferences, pushing them to adjust the content of their campaign messages accordingly.

5.1 Corporate donations and campaign quality

Corporate donations are financial resources that may enable candidates to run a better campaign. In the literature, campaign expenditures are often considered to be a means of increasing the amount of information that voters possess on candidates' policy positions and attributes (Baron, 1994; Coate, 2004b,a; Lenz, 2009; Peterson, 2009); through the organization of meetings, the distribution of leaflets or, in our context, through the provision of more details in campaign manifestos.[53]

The findings presented in Section 4.1 suggest that, indeed, donations increase the originality of manifestos' content – which serves as a proxy for the overall quality of candidate communication. However, this link is relatively weak, and certainly not as strong as the link between donations and attention to local issues. While local references may reflect a greater ability to run a targeted campaign – perhaps through increased resources, better research and better communication staff – the differential effects on local prevalence and overall originality suggest that the effect of donations on electoral discourse does not correspond entirely to an increase in campaign quality.[54]

Furthermore, if the impact of corporate donation reflected the effect of increased campaign revenue, our estimated effect should not be specific to *corporate* donations. Instead, we should see a similar impact on local prevalence for all sources of campaign funding, regardless of where the money comes from. To test for the latter hypothesis, we can only rely on correlations, as the 1995 ban applies exclusively to corporate donations. Importantly, other sources of revenue may be endogenously determined by the amount of corporate donations

---

[53]We do not have information on what candidates spend their electoral resources on. Unfortunately, this information does not exist at the level of the candidate for French data.

[54]In addition, we do not find any significant impact of corporate donations on an alternative – although imperfect – measure of manifesto quality: the total number of words it contains.

a candidate receives.[55] Nonetheless, column 1 of Table 5.1 – in which we estimate equation (2.2) including other sources of revenue as explanatory variables – provides suggestive evidence that the positive and large impact on the local index is exclusive to corporate donations: the correlation with individual donations is negative, while coefficients on the amount of personal contributions and party contributions are positive but 3 to 7 times smaller, and not significant at any conventional level.

We also find evidence that, among corporate donors, the identity of the donor matters. In columns 2 to 4, we break down the amount of corporate donations into donations made by different types of donors – as described in Section 2.2. Column 2 suggests that the effect of corporate donations on the prevalence of local-vs-national references is larger when donations are made by small donors – donors who make a single contribution – as opposed to multiple donors. Appendix Table E.14 shows that, although smaller in magnitude, this discrepancy remains when defining small donors as those making up to two or three donations (columns 1 and 2), and that it is particularly strong when defining small donors as those making up to five donations, while multiple donors are large donors who make more than five donations and whose estimated impact on the local index is much smaller and not significant (column 3).

Column 3 of Table 5.1 further shows that the estimated impact on the local index is driven both by these small donors and, among larger donors, by those who only give to candidates who are in the same district (i.e. local donors), while the effect of donations from donors who give across districts is not significant. Next, we decompose the amount of donations made by multiple donors between donors who only give to left-wing or right-wing parties, and non-partisan donors who make contributions across the political spectrum (column 4). Our results suggest that the prevalence of local references increases with contributions from these non-partisan donors exclusively, while donations from left-only or right-only donors have a *negative* (although insignificant) impact on the local index.[56]

Overall, these heterogeneous effects depending on both the source of funding and the type of corporate donor rule out a pure "resource effect" as the main mechanism behind our results: the impact of donations on candidates' propensity to advertise their local presence cannot be solely explained by the fact that corporate donations increase total campaign revenue.

---

[55]We further study and discuss possible substitution effects between corporate donations and other sources of campaign revenue in Section 6.3.

[56]Appendix Tables E.15 and E.16 show similar patterns if we consider either the frequency of local references or the frequency of national references as outcomes instead (although estimates are mostly insignificant for the latter outcome).

Table 5.1: Heterogeneity by sources of funding and type of donors

	Local index			
	(1)	(2)	(3)	(4)
Corporate donations	0.183***			
	(0.033)			
Individual donations	-0.000			
	(0.032)			
Personnal contributions	0.026			
	(0.021)			
Party contributions	0.051			
	(0.036)			
Donations from small donors		0.082**	0.079**	0.084**
		(0.038)	(0.037)	(0.038)
Donations from multiple donors		0.051*		
		(0.030)		
Multiple donors: multi-districts			0.039	
			(0.030)	
Multiple donors: single-district			0.076**	
			(0.035)	
Multiple donors: left-only				-0.019
				(0.036)
Multiple donors: right-only				-0.002
				(0.032)
Multiple donors: non-partisan				0.062**
				(0.028)
Observations	2602	2602	2602	2602
Mean outcome after ban	-0.658	-0.658	-0.658	-0.658
R2-Within	0.032	0.022	0.026	0.023

Notes: In column 1, revenue from each source of campaign funding (per voter) are divided by their respective standard deviation in 1993. In columns 2 through 4, the amount of corporate donations per voter received by each candidate is broken down into several categories depending on which type of donor they are from. In column 1, small donors make one single donation and multiple donors make donations to multiple candidates. In column 2, donors having made more than one donation are split between multi-district (donations were made in different districts) and single-district donors (all donations being made in the same district). In column 3, we separate multiple donors between donors who give only to left-wing or right-wing parties, or to both sides of the political spectrum. Other notes as in Table 4.1, column 1.

5.2 Service-induced donations and "quid-pro-quo" effect

We now discuss whether our empirical findings are driven by the "quid-pro-quo effect" of service-induced contributions. Private firms may contribute to politicians' campaigns

in exchange for economic favors or policy benefits.[57] Indeed, François and Sauger (2006) argue that French corporate donors involved in the 1993 campaign donated to seek access to power and the associated benefits – although our analysis of the district-level determinants of donations and the limited role of local economic activity calls this conclusion into question (see Section 3.1). In this paper, we do not claim to prove (or disprove) the existence of quid-pro-quo relationships between politicians and donors. What we discuss, instead, is whether such relationships – if they indeed exist – shape politicians' electoral discourse.

Section 5.1 provides some evidence consistent with a quid-pro-quo effect on campaign communication: not only do corporate donations matter, but the characteristics of the corporate donors play a role *per se*. In particular, the effect on local references is driven by small and local, non-partisan donors – who may be hoping for local economic returns. To further investigate whether different corporate donors affect differently the content of campaign manifestos, we investigate the impact of corporate donations depending on the firm's sector of activity. Appendix Table E.17 presents the results. Unsurprisingly, the estimated impact on the local index is much larger for corporate donors of an "unknown" sector (column 1), as these donors are typically small firms whose sector of activity we were not able to identify – i.e. the small donors from Table5.1. We also see a larger effect on the local index and even more so on the frequency of local references (column 2) for donors in the environmental sector.[58] These companies are typically specialized in water exploitation and distribution or waste management – activities that rely heavily on public contracts from local governments like departmental or municipal councils. Although representatives in the National Assembly cannot play a direct role in the allocation of such local contracts, advertising their local presence in their campaign communication may reflect candidates' intention to exert influence and exploit their local connections in favor of their corporate donors.[59]

To further discuss whether our empirical findings are driven by a quid-pro-quo effect, we consider two possible types of connections between politicians and donors. First, candidates may use their manifesto *ex ante* to persuade donors to contribute to their campaign by promising particularized benefits in return. As manifestos are distributed at the very end of the electoral season, it is unlikely that candidates use their manifesto as a short-run

---

[57]Note that, despite the relatively small policy autonomy of French MPs – whose positions on bill proposals are mostly contrained by party discipline (Sauger, 2009) – politicians who benefited from corporate donations may exert influence on their colleagues and party leaders to move the party line toward their donors' preferred policy platform.

[58]This sector makes the largest mean donations, and is represented by large firms such as *SPIE* in the climate engineering and energy sector or *SAUR* and the *Société des eaux* in the water supply sector.

[59]For a similar argument on political connections and French representatives' influence at the local level, see Delatte et al. (2020).

fundraising tool.[60] Second, and more realistically, politicians may "pay back" their donors after the election – and secure any future campaign contributions from them as well – so that adjusting her campaign communication is the first step in a longer-term shift in a candidate's political agenda. In this context, we expect donations to affect elected MPs' behavior and rhetoric once they are in office – and arguably more so than they affected their campaign messages. We evaluate the impact of donations on MPs' discourse once elected and their attention to local issues, by estimating equation (2.2) on the sub-sample of elected representatives.[61]

Legislative activity We first consider the overall number of written questions to the government issued by representatives during their mandate, which are typically used by politicians to voice their constituents' concerns.[62] Column 1 of Table 5.2, Panel (a), shows a negative (although insignificant) effect of corporate donations on this outcome, suggesting that donations do not induce elected politicians to raise more of these local concerns once in office. In column 2, we construct the local index described in Section 2.5 using the content of all written questions (aggregated at the representative level) and find that corporate donations do have a significant and positive impact on the prevalence of local references over national ones (8.4% of a standard deviation). However, columns 3 and 4 show that this positive effect is driven by a large negative impact on the frequency of national references (-0.5 percentage point), while the impact on the frequency of local references is also negative (-0.2 percentage point). This pattern differs substantially from the estimated impact on local references in campaign manifestos (Table 4.1, columns 1-3).

In Table 5.2, Panel (b), we estimate the impact of corporate donations on representatives' debate interventions. None of the estimates are statistically significant (possibly due to smaller sample sizes) but their directions and magnitudes suggest an effect opposite to what was observed during the campaign: donations tend to reduce the frequency of local references while increasing the frequency of national references. These legislative debates tend to examine key issues of national politics (unlike written questions), especially the most heated

---

[60]Candidates who anticipate running in a second election round may still try to secure additional contributions for the few days of campaigning leading to the runoff, but given the short period of time between the first and second rounds, this seems quite unlikely. In addition, we do not find any significant impact of corporate donations on the number of references to campaign financing in candidate manifestos, ruling out a reverse causality interpretation in which candidates attract more corporate donations because they explicitly called for such donations in their campaign communication *ex ante*.

[61]Our sample is restricted to representatives who were elected twice. In Appendix Table E.18, we re-estimate equation (2.2) on this sub-sample of elected politicians. Despite the much lower number of observations, the estimates we obtain are consistent with those presented in Table 4.1.

[62]French MPs are supposed to represent the general interest, not to defend the specific interests of their constituency, but these written questions still allow them to engage with some local issues and show their responsiveness to their constituents' needs.

debates that are more visible to the public. In Appendix Table E.19, we decompose the overall effect from Table 5.2 between the effect of donations on interventions during low-visibility debates (proxied as debates with few interventions) vs. high-visibility debates (those with many interventions). Interestingly, estimates from Panel (a) suggest that the effect of corporate donations on low-visibility debates is similar to their effect on written questions, and symmetrically opposed to their effect on high-visibility debates. It suggests that donations may influence elected politicians' discourse differently, depending on the likelihood that their words become public knowledge or not. However, this evidence is only suggestive as none of the coefficients in Appendix Table E.19 are statistically significant.

Overall, we cannot conclude that the impact of donations on candidates' local advertising in their campaign communication persists in their communication once elected. Appendix Table E.20 yields similar conclusions about the prevalence of different policy topics: while corporate donations significantly increase the prevalence of economic issues and decrease the prevalence of social issues in campaign manifestos (Table 4.2), they have no such significant effect on the prevalence of different policy topics in either written questions or debate interventions – and even decrease the prevalence of economic issues in debate interventions.

Non-elected candidates The results presented in Table 5.2 do not rule out *all* forms of political payback to corporate donors – such as the under-the-table influence politicians can exert on local politics, which we do not observe. In addition, this analysis is limited to elected politicians, while non-elected candidates may also adapt their longer-term political agenda to serve their donors' interests – especially if they hold other electoral mandates. The strong rhetorical response from non-mainstream candidates (Table 4.3) may reflect these candidates' intention to strengthen their local ties and pull more strings for their donors in the future, even though their chances of election are extremely weak – and possibly to push for long-run changes in the policy platform of their party to better match corporate interests. However, given that their influence is likely limited, the fact that niche and independent candidates are those responding the most to corporate donations in their discourse suggests that this rhetorical response does *not* reflect credible promises of particularized benefits made to donors.

In conclusion, while we are confident that the impact of corporate donations on electoral discourse is not driven by a pure resource effect (Section 5.1), the evidence of a quid-pro-quo effect is mixed. We cannot rule out that service-induced donations play a role in shaping campaign messages, but a third mechanism is likely contributing to the impact of corporate donations on electoral discourse and the increased prevalence of local references.

Table 5.2: Impact of corporate donations on legislative activity and discourse

(a) Written questions to the government

	Number of questions	Local index	Local references	National references
	(1)	(2)	(3)	(4)
Corporate donations	-4.390	0.084*	-0.017*	-0.053**
	(6.413)	(0.046)	(0.010)	(0.023)
Observations	416	416	416	416
Mean outcome	134.724	-0.781	0.187	0.598
R2-Within	0.028	0.047	0.063	0.045

(b) Debate interventions

	Number of interventions	Local index	Local references	National references
	(1)	(2)	(3)	(4)
Corporate donations	1.851	-0.063	-0.033	0.085
	(3.379)	(0.049)	(0.025)	(0.098)
Observations	356	354	354	354
Mean outcome	40.781	-1.591	0.235	3.470
R2-Within	0.042	0.023	0.011	0.018

Notes: We use one observation per elected representative per year. The sample includes all representatives elected both in 1993 and 1997. It is further restricted to candidates who issued written questions during their mandate (Table 5.2a), and those who intervened during legislative debates and whose intervention content is non-empty after text pre-processing (Table 5.2b). Other notes as in Table 4.1.

5.3 Expressive donations and electoral effect

Corporate donations may be driven by donors' preference for different politicians, and their expressive support for their preferred candidates (Ensley, 2009; Bouton et al., 2018). Fundraising activities and contact with certain corporate donors may affect candidates' perception of the issues that matter to their constituents, and induce them to adjust their campaign communication messages accordingly – i.e. an "electoral effect".

This interpretation is consistent with our different results. Receiving donations from small and local corporate donors – as well as donors who rely on local public contracts such as the environmental sector – may increase the salience of local issues, leading candidates to increase the prevalence of local references in their campaign communication (Tables 4.1 and 5.1). For instance, securing donations from a local water treatment plant may draw to the candidate's attention to the issue of water quality in the district, or the need to modernize some local infrastructures in order to protect the local environment. She will then advertise her local presence during the campaign and show voters that she is aware of these concerns.

Corporate donations may also increase the salience of economic issues – like construction and amenities – leading candidates to refer more often to such policy topics as well (Table 4.2).[63]

Niche and independent candidates may have a less precise prior belief of which issues voters care about, as compared to candidates of mainstream parties. They may therefore react more strongly to any signal about their electorate's preferences, including the signal provided by the corporate donors they make connections with.[64] As a result, marginal candidates respond more strongly to corporate donations, both in their propensity to advertise their local presence (Table 4.3, column 1) and in the choice of policy topics they refer to (Appendix Table E.12). By changing substantially their beliefs regarding voters' preferences, fundraising activities and contact with private donors may induce extreme candidates to "fall in line" and adopt an electoral discourse closer to their mainstream competitors – as further suggested by the moderation effect among Green and far-right candidates (Table 4.3, column 4).

Limitations   A first limitation of this interpretation is the small correlation between donations from individuals and the local index reported in column 1 of Table 5.1. Although this estimate is unlikely to reflect a causal relationship,[65] it suggests that receiving contributions from individuals showing their support does not incite candidates to advertise their local presence as local corporate donors do. A possible explanation for this discrepancy is the relative anonymity of individual donations: politicians are unlikely to recognize the name of every constituent sending them a check, but much more likely to know every local firm among their contributors. Furthermore, while we do not know during our period of interest the average value of the donations received from individuals (our data include information on the total amounts received but not on the number of individual donors), we know that they are capped at a lower level than corporate donations, and the anecdotal evidence we gathered from collecting novel data on contributions made in 1988 indicates that the average corporate donation is much larger than the average individual donation (see Appendix B). Candidates are also more likely to be in direct contact with local firm owners and know them personally. Hence receiving contributions from the water treatment plant may raise the salience of local environmental issues in a candidate's mind, while contributions from unknown individual

---

[63]We note here that donations from more widespread and more partisan corporate donors may also affect candidates' beliefs on the electorate's preferences, but in ways that can hardly be addressed in their individual manifesto, e.g. preferences on national policies that are decided at the national party level.

[64]Their discourse may also be more malleable, either because they have less party discipline to abide by or because their party is unestablished enough to be skirted around.

[65]We provide evidence that individual donations are endogenously affected by the ban on corporate donations in Section 6.3.

donors who care about the environment but whose preferences are unobserved does not.

A second limitation is that, as shown in columns 2-4 of Table 5.1, the positive impact of corporate donations on local prevalence is driven by small donors who give to a single candidate, but also by local donors who give to multiple candidates running in the same district, and multiple donors who give both to left-wing and right-wing politicians. These results may seem at odds with the argument that our empirical findings are driven by expressive donations, as these multiple donors are unlikely to have strong preferences over the different party platforms in the race. We argue that corporate donors who support several campaigns in a district or give indifferently to candidates of different parties may still contribute expressively to support politicians they like, not necessarily based on their party affiliation and policy platforms, but based on these politicians' attributes, social ties and friendships – either pre-dating the electoral season or established during fundraising activities.

Implication   Under the interpretation that corporate donations affect political discourse through an electoral effect, their minimal impact on discourse once a candidate is elected (Table 5.2) may reflect the existence of cheap talk: candidates adjust their campaign communication strategically to persuade voters, but these adjustments do not necessarily reflect promises to be kept once in office. Voters may give more weight and pay better attention to campaign communication than to other forms of communication – e.g. interventions from their elected representative – as they might expect politicians to *act* rather than to *talk* for them once elected. Hence candidates may be particularly responsive to their (perceived) voters' preferences in their campaign messages, but less so once the election is over. Alternatively, changes in candidates' beliefs during the campaign may be short-lived and quickly fade away. Hence our findings suggest that the influence of corporate donors may not only determine the type of information provided to voters before they cast their vote, but also the quality and reliability of this information.

6   Robustness checks and discussion

We now discuss the validity of our main result – the positive impact of corporate donations on the prevalence of local references – and provide a series of robustness checks. We discuss the main tests in the core of the paper, and report the results in Appendix Tables E.21 through E.26.

6.1 Alternative specifications

Columns 1 through 5 of Appendix Table E.21 show that the effect of corporate donations on the prevalence of local references over national ones is robust to clustering standard errors at a broader geographical level, and to using different definitions of our treatment variable (i.e. corporate donations), as described in Appendix C.

We also show that our results are robust to estimating the sample average treatment effect of corporate donations with a nearest-neighbor matching estimation (Abadie and Imbens, 2006). The idea is to match the 1997 candidates who received corporate donations in 1993 with candidates who did not but are "similar" on all other observable dimensions. Specifically, we match observations with replacement on political parties, other candidate-level controls (gender, re-running, incumbency status and holding other political mandates), and a set of district covariates as described in Section 3. In all specifications, we estimate the bias-corrected treatment effect of Abadie and Imbens (2011).

Table 6.1 presents our results. Consistently with the findings of Table 4.1, we show that the reliance on corporate donations in 1993 is associated with a *decrease* in the prevalence of local references over national ones in manifestos issued between the 1993 and 1997 legislative elections, as compared to manifestos published by otherwise "similar" candidates. In other words, reliance on corporate donations is associated with a *higher* local index (columns 1 and 2): although only marginally significant at the 10% level, the estimated effect is comparable in size, and even larger than our main estimate from column 1 of Table 4.1 – whether we match on candidate-level characteristics (column 1 of 6.1) or both candidate- and district-level characteristics (column 2). Similarly, the estimated effects on the frequency of local references and national references (respectively) are of the same sign and of the same magnitude as in Table 4.1, columns 2 and 3, when matching on all covariates (columns 4 and 6 of Table 6.1).

6.2 Parallel trends

Our difference-in-differences approach relies on the assumption that, among candidates who ran both before and after the ban, trends in campaign communication are uncorrelated with the amount of corporate donations received in 1993. To ensure its validity, our main specification (equation (2.2)) already controls both for the time-varying factors found to be associated with corporate donations in Section 3.1, and for any party-specific time trend in the prevalence of local references. Column 6 of Table E.21 shows that our estimated effect of corporate donations on the local index is robust to adding district×year fixed effects and thus controlling for district-specific time trends as well. Column 7 further shows that

Table 6.1: Impact of corporate donations on local prevalence: Nearest-neighbor matching estimation

	Local index		Local references		National references	
	(1)	(2)	(3)	(4)	(5)	(6)
ATE						
Any corporate donation	-0.175	-0.218*	-0.025	-0.316*	0.416*	0.200
	(0.143)	(0.129)	(0.206)	(0.185)	(0.246)	(0.232)
Match on candidate characteristics	✓	✓	✓	✓	✓	✓
Match on district characteristics		✓		✓		✓
Observations	1,301	1,301	1,301	1,301	1,301	1,301

Notes: Robust standard errors are shown in parentheses (***, **, * indicate significance at 1, 5, and 10 percent, respectively). We use one observation per candidate in 1997. The models are estimated using nearest-neighbor matching estimators. All specifications match on political parties (exact), classified in seven different categories: Communist party, Green party, Socialist party, right-wing party, far-right party, other parties and independent candidates. In odd columns, specifications also match on candidate-level characteristics from Figure 3.1. In even columns, specifications further match on district-level characteristics from Figure D.15. Estimates are bias-adjusted.

controlling for a full set of time-varying district-level controls, including economic indicators from firms and municipalities' finances, yields a similar result.[66] Finally, column 8 shows that the estimated impact of corporate donations on the local index is slightly smaller (14% of a standard deviation) but still large and statistically significant at the 1% level when interacting candidate controls (both present and past) with the year fixed effects, hence controlling for any differential trends across candidates with different predicted levels of corporate donations – based on their observable characteristics.[67]

A further test for the validity of our identification strategy is to show that trends in the prevalence of local references were uncorrelated with corporate donations before they were banned. However, in our context, the series of campaign finance reforms preceding the 1993 election makes such a test unreliable. Indeed, corporate donations were legalized right before the 1988 election, but we do not systematically observe which candidates may have benefited from them and to which amount – although anecdotal evidence from our ongoing data collection suggests that 1988 donations were negligible compared to 1993.[68] Hence, we cannot rule out that candidates receiving donations in 1993 had already experienced a

---

[66]This specification also controls for district-level characteristics of candidates, including the number of candidates from each party, the share of female candidates, incumbents, mayors, re-runners and candidates holding other electoral mandates. We also control for the number of registered voters and the district-level spending limit.

[67]Not all controls included in equation (2.2) are available for the 1988 elections, so our set of past controls includes indicator variables for being female and for being the incumbent. We also include categorical variables indicating if these past controls are missing, and their interaction with the year fixed effects.

[68]More details can be found in Appendix B and Appendix Figure D.19.

"treatment effect" between 1988 and 1993, if they did not receive any corporate donation in 1988 but received some for the first time in 1993. We can hardly test for any correlation between corporate donations received in 1993 and trends in communication before the 1988 reform either. The 1986 election followed a different electoral rule with a list system at the department level, so there was no candidate-level manifesto issued in that electoral year. Moreover, the 1981 election was followed by a nationwide redistricting that prevents us from linking candidates in 1988 to past district-level outcomes in 1981. Therefore, we propose a less conventional approach to show that our "treatment" (receiving corporate donations) is uncorrelated with pre-trends in outcome: we construct mean trends in manifestos' local index at the party×department level, over elections that were held since 1967. We restrict the analysis to candidates from the Communist, Socialist and right-wing parties, which span the whole period. Appendix Figure D.18 shows that, once controlling for the determinants described in Section 3, none of these pre-trends in local index is significantly correlated with the amount of corporate donations received in 1993. This suggests that candidates who benefited from corporate contributions in 1993 did *not* tend to run in departments where their party was already increasing the prevalence of local references in electoral discourse before the shocks on campaign financing occurred.

Finally, as our difference-in-differences approach relies on the inclusion of candidate fixed effects to control for the endogenous allocation of corporate donations among candidates, it mechanically restricts the sample to politicians who run both in 1993 and 1997. We discuss the threat of sample selection bias in Appendix C and provide evidence against it in column 9 of Appendix Table E.21. Overall, we are confident that our empirical strategy captures the causal impact of corporate donations on campaign communication, as opposed to the effect of confounding trends or changes in sample composition. We now discuss whether this overall impact could be driven by substitution effects with other sources of campaign revenues following the national ban.

6.3 Substitution effects

Figure 6.1 displays the composition of total revenue in 1993 and 1997 (both measured in constant euros) for candidates who received corporate donations in 1993 and for those who did not. Naturally, the amount of corporate donations drops to zero in 1997, leading to a substantial decrease in total revenues for candidates who received such donations in 1993. However, the drop in revenues for these candidates is not equal to the drop in corporate donations: while party contributions also decreased slightly between 1993 and 1997, indi-vidual donations remained roughly constant and personal contributions increased by a large

149

amount. Party contributions also decreased while personal contributions increased among candidates who did not receive any corporate donation in 1993 – leading to an overall increase in revenues for these politicians.

To further investigate potential substitution effects between corporate donations and other sources of campaign funding – while controlling for overall time trends in campaign financing, we estimate equation (2.2) using total revenues as the outcome of interest, as well as each source of revenue separately. Appendix Table E.25 shows that a one-euro-per-voter increase in corporate donations increases total revenue by 0.74 euro per voter (column 1). Conversely, it decreases the amount of individual donations by 0.05 euro per voter (column 2), the amount of party contributions by 0.14 euro per voter (column 3) and the amount of personal contributions by 0.11 euro per voter (column 4). All estimates are significant at the 1% level. The same patterns are visible when estimating the impact of increasing the share of corporate donations in total revenue on the share of revenues coming from each of the other sources (Appendix Table E.26). These results confirm that the loss of corporate donations was partly compensated by substitution effects.
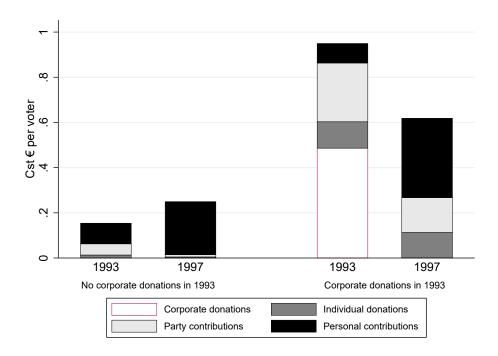
Interestingly, while we could have expected a larger substitution with individual donations – as firm owners and employees may have contributed to 1997 campaigns as individuals instead of legal entities – we find a larger effect for personal and party contributions. It suggests that candidates are able to mobilize their own resources to make up for lost revenue, and that parties act as a compensating mechanism when their candidates are hit by negative revenue shocks.[69] The latter result suggests that the limitation (or the broadening) of external sources of campaign financing may reinforce (or alter) politicians' loyalty to their party organizations, and strengthen (or loosen) parties' grip on electoral competition and representation (Katz and Mair, 1995).[70]

The existence of such substitution effects may affect the interpretation of our results and lead us to underestimate the "true" impact of donations. Although for a different purpose, we have already shown the robustness of our results to controlling for the other sources of revenue (party contributions, personal contributions and individual donations) in column 1 of Table 5.1.[71] Hence we are confident that our estimates capture the impact of receiving

---

[69]Corporate donations to parties were also banned in 1995, alleviating the concern that firms may still funnel money to specific candidates through contributions to their endorsing party. We also note that the limited substitution effect with individual donations may be partly due to the fact that such donations are capped at a lower level than corporate ones.

[70]We also note that, while the focus of this article is on manifestos produced at the local level by candidates, our results may also have implications for the ideological evolution of political parties as advertised in their manifestos (Eder et al., 2017).

[71]While these are "bad" controls, as they are endogenously determined by the change in corporate donations, this specification is informative of how the ban on corporate donations can affect campaign communication when controlling for possible substitution effects.

Figure 6.1: Revenue composition in 1993 and 1997

more corporate donations, not the impact of receiving *fewer* party contributions – or making fewer personal contributions.

Conversely, as we do not observe a *full* substitution effect between corporate donations and other sources of revenue (Figure 6.1 and column 1 of Appendix Table E.25), measuring the impact of the ban on these donations corresponds of course to measuring the impact of receiving money in 1993, no matter its source. Column 9 of Appendix Table E.21 shows that, while smaller in magnitude (.01), an increase in the share of corporate donations in total revenue significantly increases the prevalence of local references over national ones in campaign communication. It suggests that candidates do not only advertise their local presence more often when they receive more corporate donations, but also when such donations account for a larger share of their total revenues. This result echoes our discussion in Section 5 and provides additional evidence against a pure resource effect as the main driver for the impact of corporate donations on campaign communication.

7   Conclusion

This paper uses a novel dataset that combines information on donations received by candidates running for parliamentary seats and the individual campaign manifestos they issued before the elections to study the influence of money on political discourse and ideology. We exploit an historical shock on corporate donations and use a difference-in-differences approach to estimate the causal impact of donations on the content of campaign communication.

We show that donations encourage candidates to advertise their local presence in their campaign communication, at the expense of national politics. This effect is primarily driven by local donors, who contribute to several campaigns in the same district, and by non-partisan donors, who give to candidates from both left and right. Corporate donations do not shift electoral discourse toward one ideological side or the other, but push candidates from niche parties to moderate their rhetoric, to insist substantially more on local references, and to shift their discourse toward economic issues. However, we do not find any significant impact of corporate donations on legislative activity and discourse once a candidate is elected, suggesting that a quid-pro-quo effect is not the main driver of candidates' rhetorical response to donations. Instead, we argue that receiving donations from small and local corporate donors affects politicians' perception of their voters' concerns, leading them to modify their electoral discourse in favor of more local and economic references in order to win votes – without necessarily affecting their political agenda and the issues they work on once elected.

These findings shed new light on the influence of money in politics, providing evidence that campaign contributions and the nature of the donors affect what candidates talk about during electoral campaigns. While they draw from French data and a policy reform enacted in the 1990s, we argue that they are still of relevance today and hold lessons for other countries. First, in most democracies and parliamentary systems, politicians need to produce local-level communication when running for elections, through manifestos or other advertising tools. In the French context, candidate manifestos are still a primary method of communication for French politicians to address voters, and for citizens to learn about candidates.[72] Hence we expect candidates to respond to campaign contributions in the same way as they did in the 1990s. Besides, while the focus of this article is on manifestos produced at the local level by individual candidates, political parties also both strongly rely on manifestos and receive political donations. Our results may thus have implications for the understanding of the ideological evolution of political parties as advertised in their manifestos as well (Eder et al., 2017).

---

[72]In the 2017 elections for example, manifestos were used as often as social media by citizens willing to acquire information on candidates (OpinionWay, 2017), and 24% of citizens declared that they counted manifestos among the three most important ways of gathering information about candidates.

Second, bans on corporate donations have grown increasingly common over the last decade (e.g. Lithuania in 2012, Spain in 2014 or Brazil in 2015). While the focus of the campaign finance literature, both empirical and theoretical, has mainly been on large donors, we highlight the importance of considering small donors as well when evaluating campaign finance regulations (Bouton et al., 2018). On the one hand, our results suggest that banning corporate donations – and thus preventing big corporations but also small and local donors from contributing – may shift electoral discourse away from local campaigning and encourage candidates outside mainstream parties to use more polarized rhetoric – with possible longer-run consequences for the type of political discourse found in the media as well. On the other hand, campaign contributions may push voters away from their pre-advertising dispositions – not necessarily because receiving more donations allows unpopular candidates to run *more* advertising and gain prestige that way (Martin, 2014), but because receiving contributions from corporate donors influences the *content* of campaign advertising and the information made available to voters during the electoral season. While there is a large body of empirical literature documenting the effects of campaign expenditures on electoral outcomes, the debate is still ongoing as to *how* money influences voters. Our paper provides evidence that campaign contributions may influence citizens through their indirect effect on politicians: by affecting candidates' perceptions of voters' concerns, contributions shape the content of campaign messages distributed during the electoral season and, ultimately, the information voters use to form their vote decision.

References

Abadie, A. and Imbens, G. W. (2006). Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica*, 74(1):235–267.

Abadie, A. and Imbens, G. W. (2011). Bias-Corrected Matching Estimators for Average Treatment Effects. *Journal of Business & Economic Statistics*, 29(1):1–11.

Abramowitz, A. I. (1988). Explaining senate election outcomes. *American Political Science Review*, 82(2):385–403.

Adams, J. (2012). Causes and Electoral Consequences of Party Policy Shifts in Multiparty Elections: Theoretical Results and Empirical Evidence. *Annual Review of Political Science*, 15(1):401–419.

Ansolabehere, S., de Figueiredo, J. M., and Snyder, J. M. J. (2003). Why is There so Little Money in U.S. Politics? *Journal of Economic Perspectives*, 17(1):105–130.

Avis, E. (2020). Interest Groups, Campaign Finance and Policy Influence: Evidence from the U.S. Congress. Mimeo.

Avis, E., Ferraz, C., Finan, F., and Varjão, C. (2017). Money and Politics: The Effects of Campaign Spending Limits on Political Competition and Incumbency Advantage. Working Paper 23508, National Bureau of Economic Research.

Baltrunaite, A. (2020). Political contributions and public procurement: evidence from lithuania. *Journal of the European Economic Association*, 18(2):541–582.

Barber, M. J. (2016). Ideological Donors, Contribution Limits, and the Polarization of American Legislatures. *The Journal of Politics*, 78(1):296–310.

Baron, D. P. (1994). Electoral Competition with Informed and Uniformed Voters. *The American Political Science Review*, 88(1):33–47.

Bekkouche, Y. and Cagé, J. (2018). The Price of a Vote: Evidence from France, 1993-2014. CEPR Discussion Papers 12614, C.E.P.R. Discussion Papers.

Bekkouche, Y., Cagé, J., and Dewitte, E. (2020). The Heterogeneous Price of a Vote: Evidence from Multiparty Systems, 1993-2017. CEPR Discussion Papers 15150, C.E.P.R. Discussion Papers.

Bertrand, M., Bombardini, M., Fisman, R., Hackinen, B., and Trebbi, F. (2018a). Hall of Mirrors: Corporate Philanthropy and Strategic Advocacy. NBER Working Papers 25329, National Bureau of Economic Research, Inc.

Bertrand, M., Kramarz, F., Schoar, A., and Thesmar, D. (2018b). The Cost of Political Connections. *Review of Finance*, 22(3):849–876.

Boas, T. C., Hidalgo, F. D., and Richardson, N. P. (2014). The Spoils of Victory: Campaign Donations and Government Contracts in Brazil. *The Journal of Politics*, 76(2):415–429.

Bombardini, M. and Trebbi, F. (2011). Votes or money? Theory and evidence from the US Congress. *Journal of Public Economics*, 95(7–8):587–611.

Bonica, A. (2014). Mapping the Ideological Marketplace. *American Journal of Political Science*, 58(2):367–386.

Bouton, L., Castanheira, M., and Drazen, A. (2018). A Theory of Small Campaign Contributions. NBER Working Papers 24413, National Bureau of Economic Research, Inc.

Brollo, F. and Troiano, U. (2016). What happens when a woman wins an election? evidence from close races in brazil. *Journal of Development Economics*, 122:28–45.

Cagé, J. (2018). *Le prix de la démocratie*. Fayard (English version: The Price of Democracy, Harvard University Press, 2020).

Cagé, J. and Dewitte, E. (2018). It Takes Money to Make MPs : New Evidence from 150 Years of British Campaign Spending. techreport.

Carvalho, B. (2020). Campaign Spending on Local Elections: the More the Merrier? Working paper.

Catalinac, A. (2018). Positioning under alternative electoral systems: Evidence from japanese candidate election manifestos. *American Political Science Review*, 112(1):31–48.

Chamon, M. and Kaplan, E. (2013). The Iceberg Theory of Campaign Contributions: Political Threats and Interest Group Behavior. *American Economic Journal: Economic Policy*, 5(1):1–31.

Coate, S. (2004a). Pareto-Improving Campaign Finance Policy. *The American Economic Review*, 94(3):628–655.

Coate, S. (2004b). Political Competition with Campaign Contributions and Informative Advertising. *Journal of the European Economic Association*, 2(5):772–804.

Cook, R. D. and Others (2007). Fisher lecture: Dimension reduction in regression. *Statistical Science*, 22(1):1–26.

Cruz, C., Keefer, P., Labonne, J., and Trebbi, F. (2018). Making Policies Matter: Voter Responses to Campaign Promises. Working Paper 24785, National Bureau of Economic Research.

Delatte, A. L., Matray, A., and Pinardon Touati, N. (2020). Private credit under political influence: Evidence from france.

Eder, N., Jenny, M., and Müller, W. C. (2017). Manifesto functions: How party candidates view and use their party's central policy document. *Electoral Studies*, 45:75–87.

Ensley, M. J. (2009). Individual Campaign Contributions and Candidate Ideology. *Public Choice*, 138(1/2):221–238.

Erikson, R. S. and Palfrey, T. R. (1998). Campaign Spending and Incumbency: An Alternative Simultaneous Equations Approach. *The Journal of Politics*, 60(2):355–373.

Feltovich, N. and Giovannoni, F. (2015). Selection vs. accountability: An experimental investigation of campaign promises in a moral-hazard environment. *Journal of Public Economics*, 126:39–51.

Fisman, R. (2001). Estimating the Value of Political Connections. *The American Economic Review*, 91(4):1095–1102.

Foucault, M. and François, A. (2005). Le rendement des dépenses électorales en France. Le cas des élections législatives de 1997. *Revue économique*, 56(5):1125–1143.

Fouirnaies, A. and Hall, A. B. (2018). How Do Interest Groups Seek Access to Committees? *American Journal of Political Science*, 62(1):132–147.

Fowler, A., Garro, H., and Spenkuch, J. L. (2020). Quid pro quo? corporate returns to campaign contributions. *The Journal of Politics*, 82(3):844–858.

François, A. and Phélippeau, É. (2015). *Le financement de la vie politique: Réglementations, pratiques et effets politiques.* Armand Colin.

François, A. and Sauger, N. (2006). Groupes d'intérêt et financement de la vie politique en france. *Revue française de science politique*, 56(2):227–254.

François, A., Visser, M., and Wilner, L. (2016). Using Political Financing Reforms to Measure Campaign Spending Effects on Electoral Outcomes. CESifo Working Paper 6232.

Gentzkow, M., Shapiro, J. M., and Taddy, M. (2019). Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech. *Econometrica*, 87(4):1307–1340.

Gerber, A. (1998). Estimating the Effect of Campaign Spending on Senate Election Outcomes Using Instrumental Variables. *The American Political Science Review*, 92(2):401–411.

Gerber, A. S. (2004). Does Campaign Spending Work? *American Behavioral Scientist*, 47(5):541–574.

Gimpel, J. G., Lee, F. E., and Kaminski, J. (2006). The Political Geography of Campaign Contributions in American Politics. *The Journal of Politics*, 68(3):626–639.

Gordon, S. C., Hafer, C., and Landa, D. (2007). Consumption or Investment? On Motivations for Political Giving. *The Journal of Politics*, 69(4):1057–1072.

Green, D. P. and Krasno, J. S. (1988). Salvation for the Spendthrift Incumbent: Reestimating the Effects of Campaign Spending in House Elections. *American Journal of Political Science*, 32(4):884–907.

Grossman, G. M. and Helpman, E. (1994). Protection for Sale. *American Economic Review*, 84(4):833–850.

Gulzar, S., Rueda, M. R., and Ruiz, N. A. (2021). Do campaign contribution limits curb the influence of money in politics?

Gunlicks, A. B. (1993). *Campaign and Party Finance in North America and Western Europe*. Westview Press.

Jacobson, G. C. (1978). The Effects of Campaign Spending in Congressional Elections. *The American Political Science Review*, 72(2):469–491.

Jacobson, G. C. (2006). Campaign spending effects in U.S. Senate elections: Evidence from the National Annenberg Election Survey. *Electoral Studies*, 25(2):195–226.

Kalla, J. L. and Broockman, D. E. (2016). Campaign contributions facilitate access to congressional officials: A randomized field experiment. *American Journal of Political Science*, 60(3):545–558.

Kalla, J. L. and Broockman, D. E. (2018). The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments. *American Political Science Review*, 112(1):148–166.

Katz, R. S. and Mair, P. (1995). Changing models of party organization and party democracy: the emergence of the cartel party. *Party politics*, 1(1):5–28.

Kendall, C., Nannicini, T., and Trebbi, F. (2015). How Do Voters Respond to Information? Evidence from a Randomized Campaign. *American Economic Review*, 105(1):322–353.

Le Pennec, C. (2020). Strategic campaign communication: Evidence from 30,000 candidate manifestos. SoDa Laboratories Working Paper Series 2020-05, Monash University, SoDa Laboratories.

Lenz, G. S. (2009). Learning and Opinion Change, Not Priming: Reconsidering the Priming Hypothesis. *American Journal of Political Science*, 53(4):821–837.

Levitt, S. D. (1994). Using Repeat Challengers to Estimate the Effect of Campaign Spending on Election Outcomes in the U.S. House. *Journal of Political Economy*, 102(4):777–798.

Martin, G. J. (2014). The informational content of campaign advertising. *mimeographed, Emory.*

McCarty, N., Poole, K. T., and Rosenthal, H. (2016). *Polarized America: The Dance of Ideology and Unequal Riches.* Walras-Pareto Lectures. MIT Press.

Morton, R. and Cameron, C. (1992). Elections And The Theory Of Campaign Contributions: A Survey And Critical Analysis. *Economics and Politics*, 4(1):79–108.

OpinionWay (2017). Les Français et les programmes électoraux. Sondage OpinionWay pour Le Printemps de l'Economie.

Palda, F. and Palda, K. (1998). The Impact of Campaign Expenditures on Political Competition in the French Legislative Elections of 1993. *Public Choice*, 94(1/2):157–174.

Peterson, D. A. M. (2009). Campaign Learning and Vote Determinants. *American Journal of Political Science*, 53(2):445–460.

Pons, V. and Tricaud, C. (2019). The large effects of a small win: How past rankings shape the behavior of voters and candidates. Technical report, National Bureau of Economic Research.

Powell, E. N. and Grimmer, J. (2016). Money in exile: Campaign contributions and committee access. *The Journal of Politics*, 78(4):974–988.

Rhodes, J. H., Schaffner, B. F., and Raja, R. J. L. (2018). Detecting and Understanding Donor Strategies in Midterm Elections. *Political Research Quarterly*, 71(3):503–516.

Sauger, N. (2009). Party discipline and coalition management in the french parliament. *West European Politics*, 32(2):310–326.

Taddy, M. (2013). Multinomial Inverse Regression for Text Analysis. *Journal of the American Statistical Association*, 108(503):755–770.

Taddy, M. (2015). Distributed multinomial regression. *The Annals of Applied Statistics*, 9(3):1394–1414.

Taddy, M. (2017). One-step estimator paths for concave regularization. *Journal of Computational and Graphical Statistics*, 26(3):525–536.

Teso, E. (2020). Whar Drives US Corporate Elites' Campaign Contribution Behavior?". Working paper.

Titl, V. and Geys, B. (2019). Political donations and the allocation of public procurement contracts. *European Economic Review*, 111:443–458.

Appendices

A Data

A.1 Campaign manifestos

Campaign manifestos are a key part of the French electoral campaigns, and represent one of the three main parts of official electoral propaganda (together with ballots and election posters). Candidates are responsible for the printing of these manifestos, whose cost can be refunded by the state if they gather at least 5% of the votes during one of the two rounds of the election (Electoral law, articles R39 and L216). The mailing is taken in charge by an official local propaganda committee, if the format of the manifestos respects certain criteria. More specifically, electoral manifestos must have a maximum size of 210x297 millimeters, and a weight ranging between 60 and 80 grams per square meter (Electoral law, article R29). Furthermore, they cannot combine the three colors of the French flag (blue, white and red, article R27 of the electoral law), except if they are part of a party's emblem. If these constraints are met, the manifestos are mailed to voters, together with ballots, maximum four days before the election (for the first round), and three days before the second-round (in case of a runoff) (Electoral law, articles R34 and R38).

In a survey published before the 2017 Presidential election (OpinionWay, 2017), 24% of citizens declared that manifestos were among the three most important ways of getting information about candidates. By comparison, television was mentioned by 64% of them, online media by 26%, paper news by 18% and radio by 15%. The fact that, in 2017, candidates' manifestos were mentioned about as often as online media suggests that they are not a negligible part of the heavy campaign communication voters receive during the few weeks leading to the election. In all likelihood, this number is a lower bound for the share of voters who learnt about their candidates thanks to the manifestos over our sample period, when much fewer communication media were available to individual politicians. Of course, television was already an important medium of communication. But while TV shows, debates and ads are the prominent media for candidates who campaign at the national level - such as candidates to the Presidential elections or party leaders who advertise their national platform before the Legislative elections - it is unlikely that voters learn much about the individual candidates running in their district on TV. Conversely, individual manifestos are a prime method of communication for candidates to run their own campaign and tailor the message to the specific voters in their district.

A.2   Text as data

Text pre-processing

We pre-process the content of each corpus in our dataset (manifestos, questions to the governments, debate interventions) following standard steps from the literature. To construct our measure of local anchoring, we tokenize documents at the single word level and remove stopwords and special characters – both in the documents' content and in our local and national dictionaries. For all other measures, we also lemmatize each word[73] and restrict the vocabulary to words used by at least 0.5% and at most 50% of the documents – with the exception of the vocabulary used to estimate word loadings for different policy topics, which is restricted to words used by at least 0.1% of all written questions issued between 1988 and 1997, due to the large number of such questions (close to 200,000).

Multinomial inverse regression

We describe here the framework introduced by Taddy (2013). The frequency of word w in document j - c_{wj} - is derived from a discrete choice model over the vocabulary of size W and is assumed to follow a multinomial distribution of the form $c_{wj} \sim MN(q_{wj}, m_j)$, where m_j is the number of words in document j. To construct a document's left-right score on the left-right scale, we define the probability of document j using word w as:

$$q_{wj} = \frac{exp(\alpha_w + \phi_w D_j)}{\sum_{k=1}^{W} exp(\alpha_k + \phi_k D_j)}$$

where D_j is an indicator variable equal to one if j is issued by a right-wing candidate as opposed to a left-wing one. Non-classified and centrist candidates are excluded. ϕ_w is a word *loading* that measures sensitivity to party affiliation or the gain in utility from using this word for a right-wing candidate over a left-wing one. A sufficient reduction (Cook and Others, 2007) for j's partisanship given the observed vector of word frequencies is the following projection:

$$Z_j = \sum_{w=1}^{W} \phi_w \cdot \frac{c_{wj}}{m_j}$$

where Z_j is the left-right partisan score of document j: a negative (positive) score means that document j uses a lot of words used by other left-(right-)wing candidates but never by the other side, while a score close to zero means that document j uses either neutral words used by both sides indifferently, or a mix of polarizing words from both sides.

The parameters of interest α_w and ϕ_w are estimated through distributed multinomial regression (Taddy, 2015), where a Poisson approximation for the distribution of c_{wj} allows for faster and more efficient distributed computing. The implied negative log-likelihood for each word is proportional to:

$$l(\alpha_w, \phi_w) = \sum_{j=1}^{N} [m_j exp(\alpha_w + \phi_w D_j) - c_{wj}(\alpha_w + \phi_w D_j)]$$

---

[73]We use Spacy's French model to lemmatize our vocabulary: https://spacy.io/models/fr

Following Gentzkow et al. (2019), we control bias through penalization. In particular, we apply the gamma-lasso procedure described in Taddy (2017) so that the preferred estimator is:

$$\hat{\alpha_w}, \hat{\phi_w} = argmin[l(\alpha_w, \phi_w) + N\lambda\gamma^{-1}log(1 + \gamma|\phi_w|]$$

where N is the number of documents in the corpus, λ is a standard Lasso penalty, and γ is the penalty scale.[74] This penalized estimator shrinks noisy loadings to zero, resulting in a sparse solution that downweights the artificially high influence of rare words in the corpus. The estimation is implemented with the textir library in R.

Policy topics We follow essentially the same strategy to project manifestos onto latent topics and policy fields, using the sample of written questions to the government issued between 1988 an 1997 as training set.[75] More specifically, we define the probability of document j using word w as:

$$q_{wj} = \frac{exp(\alpha_w + \sum_{s=1}^{S} \phi_w^s D_j^s)}{\sum_{k=1}^{W} exp(\alpha_k + \sum_{s=1}^{S} \phi_k^s D_j^s)}$$

D_j^s is an indicator variable equal to one if question j is addressed to a minister about topic s. ϕ_w^s is a word *loading* that measures the lift in utility from using word w when issuing a question about topic s as opposed to targeting a non-classified ministry.[76] The sufficient reduction for the topic assignment of any document j - given the observed vector of word frequencies - is the following projection:

$$Z_j^s = \sum_{w=1}^{W} \phi_w^s \cdot \frac{c_{wj}}{m_j}$$

This quantity provides a continuous measure for the prevalence of topic s in document j. Intuitively, a document with a high positive Z^s is a document that uses many words whose loading - or predictive power - for topic s is also high. We can use the set of parameters ϕ_w^s estimated from written questions to the government to project manifestos onto each latent topic space and obtain a set of topic prevalence measures for each manifesto.

To further obtain measures of topic prevalence that are easily interpretable, we feed the set of continuous measures Z^s into a multinomial logistic regression of the form:

$$P(D_j = s) = \frac{exp(\alpha_s + \sum_{s'=1}^{S} \delta_s^{s'} Z_j^{s'})}{\sum_{s'=1}^{S} exp(\alpha_{s'} + \sum_{s'=1}^{S} \delta_{s'}^{s'} Z_j^{s'})}$$

where $P(D_j = s)$ is the probability that document j refers primarily to topic s. We fit the

---

[74]For details on the advantages of concave regularization and Gamma Lasso versus Lasso penalization, see Taddy (2017).

[75]For the training stage, we restrict the vocabulary to words used in at least 0.1% – instead of 0.5% – of all questions issued over that period, which corresponds to about 200 questions.

[76]The intercept of this model corresponds to the baseline utility of using word w when issuing a question to any non-classified minister.

model on the sample of written questions to the government, using 80% of the observations (randomly chosen) as training set and the other 20% as a test set to evalaute the out-of-sample performance of the model. We obtain 86% accuracy with 17 topics and 87% accuracy with 4 broader topics. We then use the estimated set of δ_s coefficients – as well as the manifesto projections Z^s – to assign each manifesto to a set of estimated probabilities, each indicating the likelihood that the manifesto focuses primarily on a given topic over the others.

Latent Semantinc Indexing

Following Bertrand et al. (2018a), we use Latent Semantic Indexing to construct measures of pairwise similarity between each pair of manifestos among candidates from the same party. To implement this simple bag-of-words approach, we first represent our corpus of manifestos as a document-term matrix, where each manifesto is represented as a vector of Tf-Idf weights over the pre-processed vocabulary. These weights increase with document specificity: a word with a large Tf-Idf weight is a word that is frequent in a given document but not so frequent across the whole corpus. We then apply a singular value decomposition to this large and sparse document-term matrix to reduce its dimensionality and obtain a dense matrix, where each document is represented as a vector of 200 latent dimensions.[77] We measure the cosine similarity between each pair of such dense vectors, and define the originality index as the mean (negative) similarity between a candidate manifesto and each other manifesto from the same party. This measure is further standardized by year for interpretability.

A.3   Corporate donations

Data on corporate donations to candidates in 1993 come from the reports published by the CNCCFP after the examination of candidates' account. For each candidate, we digitize the campaign accounts that include the comprehensive list of corporate donors and the amounts given. An example of the data is shown in Appendix D.1. In total, 14,770 donations were received by 1,647 candidates (so around one third of the candidates). We show descriptive statistics on these corporate donations in Appendix 2.2.

Donor identification

The first step of the cleaning consisted in creating a unique donor identifier. We retrieve the list of all donors' name as they appear in the reports and remove stopwords, and homogenize numeric characters in plain words. For national companies where the local branch was specified in the donor name, we attribute a common donor code. For instance, the firm *COLAS* gave to candidates through its subsidiaries *COLAS MEDITERRANNEE, COLAS SUD OUEST* or *COLAS MEDITERRANNEE*. To separate firms including a geographical attribute in its legal denomination from local branches, we use an algorithm to check on the website societe.com whether the company was considered as the mother entity. Yet, a certain number of firms active in 1993 have ceased activity since the election and their record is not available online. We conduct a second search using data from the INSEE (the

---

[77]The number of dimensions is chosen arbitrarily and motivated from existing research.

French national statistical institute) dataset of French firms active in 1993. At the end of the procedure, we are left with 10,470 unique donors.

As a note of caution, we cannot exclude that a firm appearing with two different names and not matched with the INSEE dataset – for instance an entity named both with an acronym and with the plain denomination, is not considered as two different donors. Yet, we conduct further manual checks that make us confident that these cases can be considered as measurement error. Further, to avoid bias stemming from this type of error, we choose to distinguish between single and multiple donors rather than considering the number of donations of each donor in the empirical analysis of Section 5. This allows us to test for the robustness of the results on the heterogeneity by defining multiple donors as entities giving more than 1, 2, or 5 donations (see Section 5 and Appendix Table E.14).

Sectors of activity

To complement our donor dataset, we look at their sector of activity. Given the format of the raw data that only provide the name of the donor without any further information or firm identifier, and the fact that the data date back to 1993, retrieving this sector is a challenging exercise. To do so, we implement different procedures: we first merge the donors with firm records from the INSEE. Second, we take advantage of the fact that firms' name are sometimes explicit about the type of activity of the donor and therefore use those to manually classify corporations.[78] At the end of the procedure, we manage to identify the sector of about half of the sample: not surprisingly, larger donors are more likely to be tied to a sector and there is a wide and significant imbalance between the average donation made by sector-identified firms and others (see Appendix Table E.4). Appendix Table E.5 shows summary statistics across sectors of activity: the most represented sector is construction, followed by retail sector, which encompasses large retail companies or smaller businesses. Discrepancies in terms of donations by sector are also to be highlighted: as shown in Appendix Figure D.4, both the number of donations per donor and the average donation amount are higher among donors from the environment/energy and the construction sectors.

A.4   Other data

Finally, we collect time-varying district-level demographic covariates. Demographic and unemployment data are from the French census. To understand the determinants of corporate donations, we build a new dataset on the revenues and annual spending in infrastructure of the French municipalities with more than 10,000 inhabitants, from the paper-format archives of the Ministry of Finances covering the 1993-1997 time period.

Our dataset also includes the annual number of firms, of employees, the total payroll, as well as the share of the employees who are part of the top 1% of the income distribution. These are from the *"Déclaration Annuelle de Données Sociales"* (DADS), a detailed French database on wages.

Other available district-level factors include the number of municipalities in the district, rural-urban status, and whether the capital of the region belongs to the district. Summary

---

[78]We use the set of firms that we successfully classified with the two procedures to refine the manual cleaning strategy.

statistics on these covariates are shown in Appendix Table E.9.

B   1988 legislative elections

Donations were first allowed with the laws passed in March 1988, and candidates at the 1988 legislative elections that took place on June 5th and 12th were thus entitled to receive contributions both from individuals and corporations. Yet, the campaign accounts of the 1988 candidates have never been studied until now, including by historians. This is due to the fact that, in the absence of a centralized regulatory agency – the *"Commission Nationale des Comptes de Campagne et des Finances Politiques"* (the French equivalent of the US FEC) was only created in 1990 – these accounts have not been validated neither assembled in the National archives (or in the archives of the Commission). A careful reading of the administrative rules in place and numerous interactions with archivists led us on the trail of the departmental archives. A number of these archives indeed store as of today the 1988 candidates' campaign accounts. However, because the identity of the individual donors has not been anonymized, the documents are still classified.

We have contacted separately the persons in charge of each of the departmental archives holding the accounts (96 departments in Metropolitan France), and asked officially for the declassification of the documents (given our approach is purely research driven). For now, we have only received one positive answer, from the departmental archives of Seine Maritime. While obviously incomplete and thus imperfect, this allows us to get a sense of the structure of the donations and expenditures at the 1988 legislative elections for 12 electoral districts, and to compare them to the 1993 and 1997 elections. Out of the 56 candidates running in the Seine Maritime department in 1988, 28 also ran in 1993. We compare their revenues and expenses during these two electoral years. Appendix Figure D.19 reports the results. It appears clearly that candidates both received and spent much less in 1988 than in 1993. It is not surprising given the possibility of receiving donations was a new opportunity, offered to the candidates only three months before the elections.

C   Robustness checks

C.1   Alternative specifications

Clustering   In our preferred specification, we cluster the standard errors at the district level. The estimates remain significant when clustering standard errors at the department level instead (Appendix Table E.21, column 1).

Measuring corporate donations   We test for the robustness of our estimates to using alternative measures of corporate donations. Column 2 of Appendix Table E.21 shows that estimating equation (2.2) with the (standardized) log of corporate donations[79] as independent variable yields an estimated impact of donations on a manifesto's local index that is slightly larger (0.18) to the point estimate from column 1 of Table 4.1. In column 3, we use an indicator variable for receiving any corporate donation as independent variable, which shows that the effect of corporate donations is even higher at the extensive margin, with an estimated effect corresponding to 24% of a standard deviation in the local index, significant at the 1% level.

Column 4 shows a smaller (0.02) but significant estimate for the effect of the number of (distinct) corporate donations received by each candidate. In column 5 we estimate a quadratic version of equation (2.2) and find that the effect of corporate donations on the prevalence of local references over national ones follows a convex pattern, indicating that the positive impact wears off as candidates receive larger and larger amounts of donations. Interestingly, these two columns suggest that receiving few important donations – rather than many – is what affects campaign communication the most. This pattern is consistent with our preferred interpretation of the results, presented in Section 5: the support of a few committed corporate donors is likely to increase the salience of certain issues and push candidates to address these issues in their campaign communication – such as local issues. Receiving a large amount of contributions but from a wide array of different donors may not provide such a clear signal of which issues constituents care about

C.2   Sample selection

Our difference-in-differences approach relies on the inclusion of candidate fixed effects. While this strategy controls for the endogenous allocation of corporate donations among candidates with different unobserved attributes – which is arguably the greatest threat to causal identification – it mechanically restricts the sample to candidates who run both in 1993 and 1997. The subsample of re-runners differs significantly from the overall sample of candidates: as shown in Appendix Table E.23, among all candidates running in 1993, those who ran again in 1997 are more likely to be men, to have already run in the past, to have won the previous election, to hold another electoral mandate and to enjoy higher campaign revenues – including corporate donations. These systematic differences may threaten the external validity of our results, as they may not apply to candidates who ran only once. It may also threaten the internal validity of our approach, if the amount of corporate donations

---

[79]More precisely we use $ln(\text{Corporate Donations}_{ipdt} + 1)$ as independent variable to account for the many zeros in the data. We then divide that quantity by its standard deviation in 1993.

received in 1993 pushes candidates of a certain type and with certain communication skills to run again in 1997. Column 1 of Appendix Table E.24 suggests that a one-standard-deviation increase in corporate donations received in 1993 raises the probability that a candidate runs again in 1997 by 3.4 percentage points – an estimate significant at the 1% level. This specification includes all candidates who run either in 1993 or in 1997. We estimate a regression model of the form of equation (2.2), where the outcome is an indicator variable equal to 1 if the candidate runs again in the next election (1997 or 2002) and where we replace candidate fixed effects with district fixed effects. Column 2 of Appendix Table E.24 also shows that corporate donations have a small but significant impact on manifesto availability, which determines whether a candidate is included in our sample as well.

To alleviate this concern of endogenous sample selection, we test for the robustness of our results to a less conservative approach, in which we replace candidate fixed effects with party×district fixed effects and include all candidates whose party is present in the same district twice – even if it was not the same candidate running in both years. This specification excludes independent candidates. Column 9 of Appendix Table E.21 shows a positive estimate of corporate donations on the local index, significant at the 1% level. Interestingly, the point estimate is smaller in magnitude (.1) as compared to column 1 of Table 4.1, suggesting that the within-party allocation of corporate donations in 1993 is biased toward individual politicians who, absent any donation, would be *less* likely to make local references in their manifesto.

D   Additional Figures to Chapter 2

ÉLECTIONS LÉGISLATIVES GÉNÉRALES DES 21 MARS ET 28 MARS 1993

AISNE (Iʳᵉ circonscription)

Plafond de dépenses : 500 000 F   Décision C.C.F.P. du : 05-11-93   Scrutin non contesté

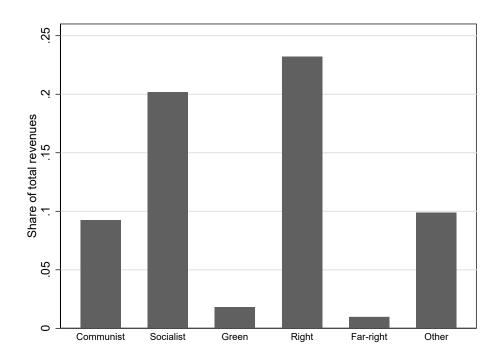| DÉPENSES				RECETTES							NOMS DES CANDIDATS	Décisions C.C.F.P.
Total déclaré	Base R. 39	Réforma-tions	Total retenu	Dons P.P.	Dons P.M.	Apport personnel net	Apport parti net	Autres	Réforma-tions	Total retenu		
18 473	0	0	18 473	0	0	0	18 473	0	0	18 473	PERNELLE Jean-Loup	A
236 465	58 501	+40 852	218 816	34 200	77 750	49 695	40 852	19 614	0	222 111	DOSIERE René	AR
98 344	36 536	0	61 808	4 350	0	57 458	0	0	0	61 808	SALECK Michel	A
392 614	59 862	0	332 752	85 750	159 800	0	200 000	55 328	0	500 878	LAMANT Jean-Claude	A
53 567	16 395	0	37 172	0	0	26 131	10 041	1 000	0	37 172	DEGEMBE Patrick	A
74 570	26 131	0	48 439	1 000	0	47 439	0	0	0	48 439	LACOMBE Dominique	A
33 173	0	+300	33 473	0	0	0	33 173	0	+300	33 473	BERDAL Michelle	HD
0	0	0	0	0	0	0	0	0	0	0	JARNO Philippe	ND

Listes des dons de personnes morales versés à partir du Iᵉʳ février 1993 (loi nᵒ 9²-122 du 29 janvier 1993)

AISNE (Iʳᵉ circonscription)

René DOSIERE

STE ANIZIENNE DE CONSTRUCTION .... 10 000 F

STE ANIZIENNE DE VIABILITE ET D'ASSAINIS-SEMENT.. 3 000 F

ENTREPRISE DE CONSTRUCTION DE LA THIE-RACHE .... 1 000 F

SARL SCOP CHAUFFAGE SANITAIRE.................... 2 000 F

CAISSE MUTUELLE D'ASSURANCES...................... 1 000 F

ARCHITECTES ASSOCIES BORDERIOUX DI LEGGE.. 7 000 F

SA BOUCARD... 2 500 F

SA EUROP ALU.. 750 F

Jean-Claude LAMANT

ETS CAILLE SA.. 1 000 F
ENTREPRISE DRAPIER.. 8 000 F
SA DU PARC.. 3 000 F
SA LE BETON ARME FERRARI................................ 5 000 F
SA BPF... 2 000 F
SARL GARAGE LESOUDARD.................................... 1 000 F
SA CHAMBRY DISTRIBUTION.................................. 15 000 F
ENTREPRISE CHEMERY .... 1 500 F
ENTREPRISE CHEMERY .... 1 000 F
SAB ENTREPRISE BOUCARD.................................... 2 500 F
SARL VITRANT... 5 000 F
GENERALE DE TRANSPORT ET D'INDUSTRIE..... 30 000 F
SA THOURAUD.. 5 000 F
S.G.S.T SAVE... 10 000 F
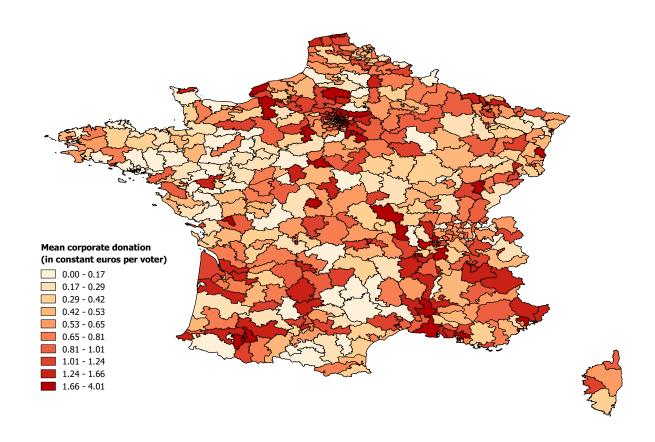
Notes: The figure provides an example of the CNCCFP's paper archives from which we have collected information on the corporate donations received by the 1993 legislative elections candidates, including the name of the corporate donors and the amount of their donation.

Figure D.1: Example of corporate donations data
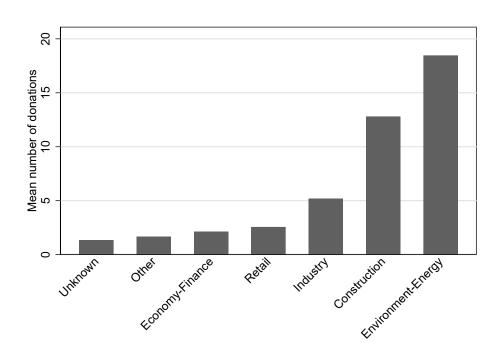
Notes: The figure shows the ratio of the mean revenues from corporate donations over the mean revenue of candidates affiliated with the party in 1993, for the five main parties. The category "Other" includes independent candidates and candidates running for a minor party without a national foothold in 1993. N=5,141.

Figure D.2: Share of mean corporate donations in mean total revenue across parties
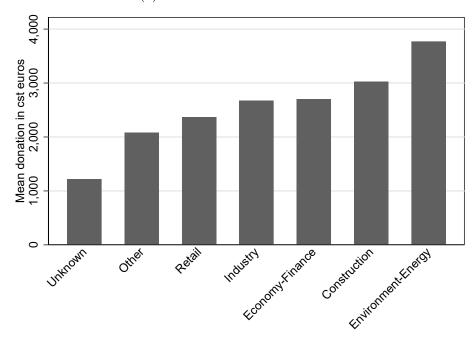
Notes: The map shows the mean value of corporate donations received by candidates running in a district in 1993, in constant euros per voter. Districts are split in deciles: districts in the lightest orange bracket belong to the 10% of districts with the lowest average of corporate donations per candidates (between 0 and .17 euro per voter on average in the district), the darkest red standing for the 10% districts with the highest value of corporate donations (between 1.66 and 4 euro per voter). N=555.

Figure D.3: Mean corporate donations in 1993

170

(a) Mean number of donations
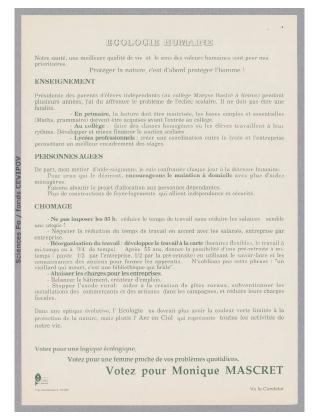
(b) Mean amount in constant euros

Notes: Figure D.4a displays the mean number of donations per donor, and Figure D.4b displays the mean donation in constant euros, by sector of activity. Sectors with less than 500 donations are grouped in the category "Other".
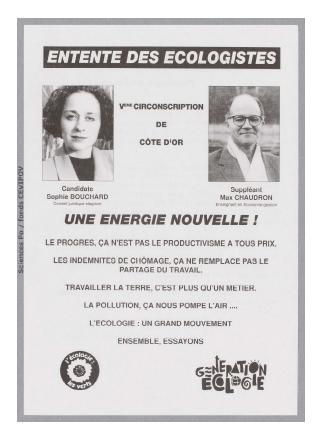
Figure D.4: Donations by sector of activity

Figure D.5: Manifesto from a Green candidate with corporate donations
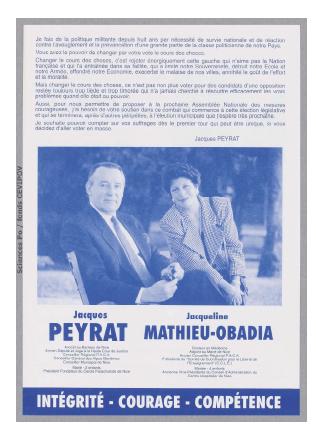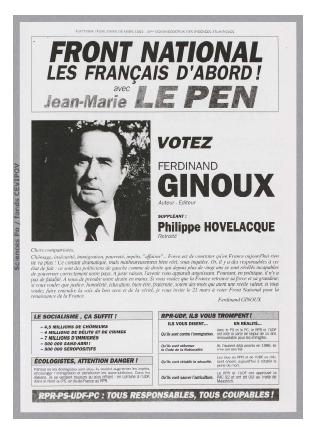
Figure D.6: Manifesto from a Green candidate with no corporate donation
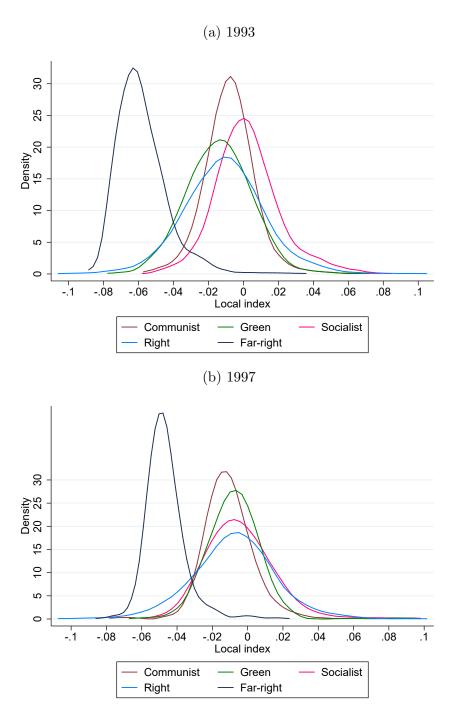
Figure D.7: Manifesto from a far-right candidate with corporate donations
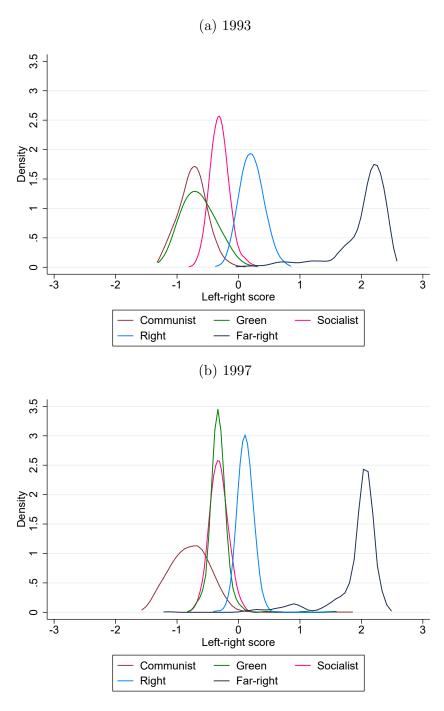
Figure D.8: Manifesto from a far-right candidate with no corporate donation

(a) 1993

(b) 1997
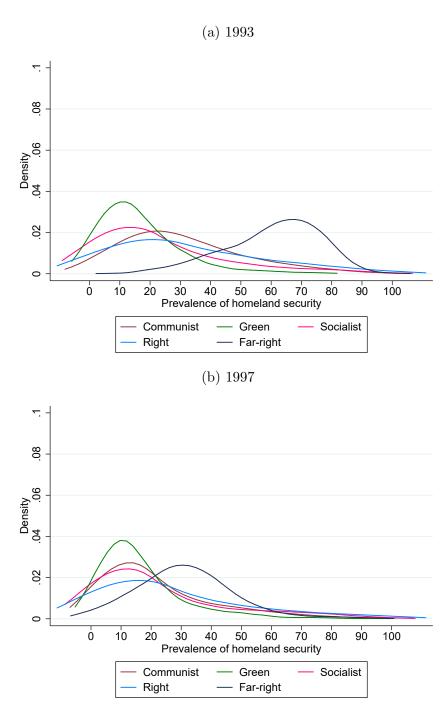
Notes: We plot – for each of the five main parties in our sample – the kernel density of manifestos' local index, which measures the prevalence of local references over national ones, in 1993 and in 1997 separately. The sample includes all candidates from the Communist party, the Green party, the Socialist party, the conservative right-wing party and the far-right party, whose first-round manifesto is available and non-empty after text pre-processing. Large outliers are excluded for visual purposes. N=2,535 and N=2,528 (resp.).

Figure D.9: Kernel density of the local index by party
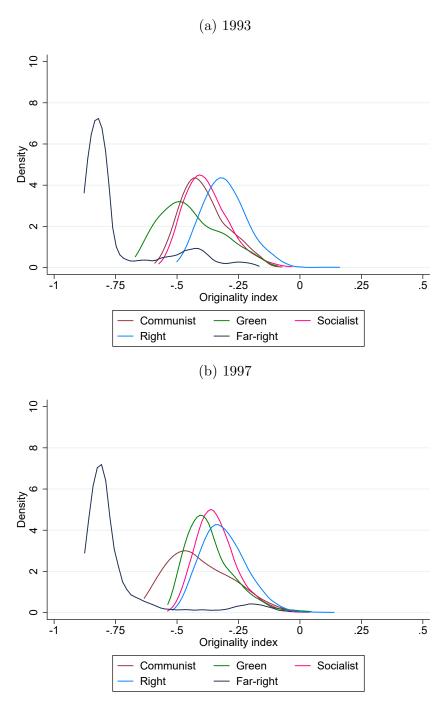
(a) 1993

(b) 1997

Notes: We plot – for each of the five main parties in our sample – the kernel density of left-right scores from manifestos (issued before the first election round), in 1993 and in 1997 separately. This score indicates the partisan leaning of each manifesto from left-wing (negative score) to right-wing (positive score), based on the words it contains. Other notes as in Appendix Figure D.9.

Figure D.10: Kernel density of left-right score by party

(a) 1993

(b) 1997

Notes: We plot – for each of the five main parties in our sample – the kernel density of homeland security prevalence in manifestos (issued before the first election round), in 1993 and in 1997 separately. The prevalence of homeland security indicates the probability (in percentage points) that the manifesto focuses primarily on homeland security issues out of 17 policy topics, based on the words it contains. Other notes as in Appendix Figure D.9.
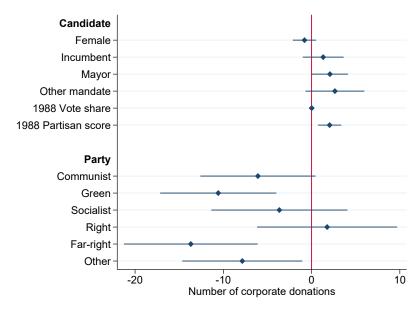
Figure D.11: Kernel density of homeland security prevalence by party
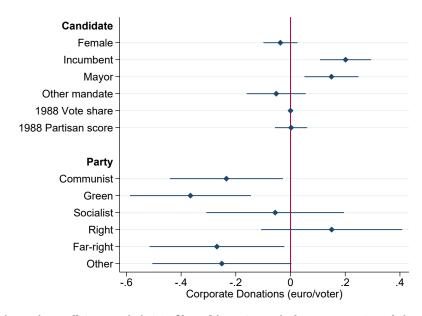
(a) 1993

(b) 1997

Notes: We plot – for each of the five main parties in our sample – the kernel density of candidate originality (issued before the first election round), in 1993 and in 1997 separately. The originality index indicates whether a manifesto is similar to (lower value) or distinct from (higher value) other manifestos from the same party. Other notes as in Appendix Figure D.9.

Figure D.12: Kernel density of candidate originality by party

(a) Number of corporate donations
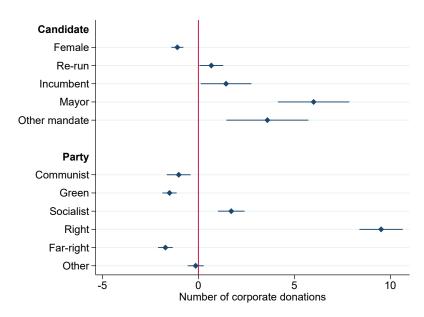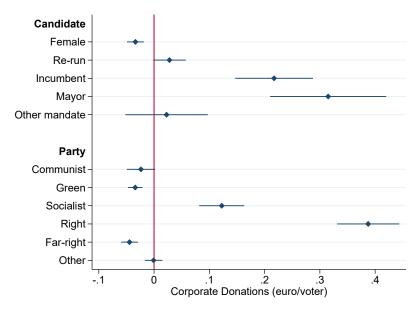
(b) Amount of corporate donations

Notes: The figure shows the coefficients and their 95% confidence intervals from a regression of the number of corporate donations (Figure D.13a) or the amount of corporate donations in constant euros per voter (Figure D.13b) received by each candidate on a set of party fixed effects (omitting independent candidates) and candidate characteristics. We use one observation per candidate in 1993. The sample is restricted to candidates who ran both in 1988 and 1993 and whose 1988 manifesto is available. Standard errors are clustered at the district level.

Figure D.13: Candidate-level determinants of corporate donations in 1993, Controlling for 1988 left-right score

(a) Number of corporate donations
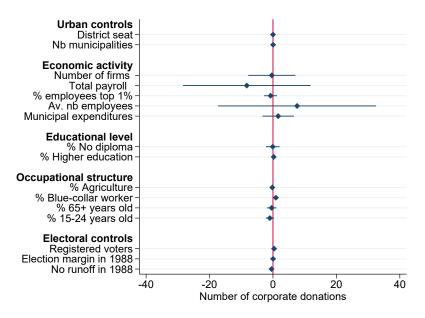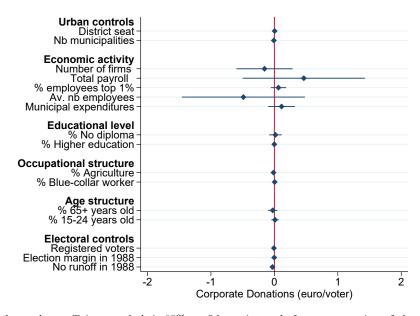
(b) Amount of corporate donations

Notes: The figure shows the coefficients and their 95% confidence intervals from a regression of the number of corporate donations (Figure D.14a) or the amount of corporate donations in constant euros per voter (Figure D.14b) received by each candidate on a set of district fixed effects, party fixed effects (omitting independent candidates), and candidate characteristics. We use one observation per candidate in 1993. Standard errors are clustered at the district level.

Figure D.14: Candidate-level determinants of corporate donations in 1993, Controlling for district fixed effects

(a) Number of corporate donations
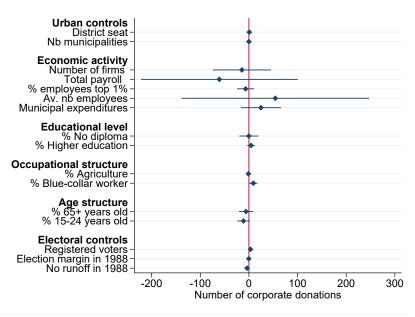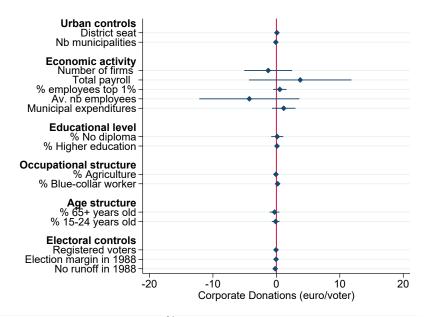
(b) Amount of corporate donations

Notes: This figure shows the coefficients and their 95% confidence intervals from a regression of the number of corporate donations (Figure D.15a) or the amount of corporate donations in constant euros per voter (Figure D.15b) received by each candidate on a set of party fixed effects, candidate characteristics, and district characteristics (estimation of equation 2.1). All explanatory variables are standardized. We use one observation per candidate in 1993. Standard errors are clustered at the district level.

Figure D.15: District-level determinants of corporate donations in 1993

(a) Number of corporate donations
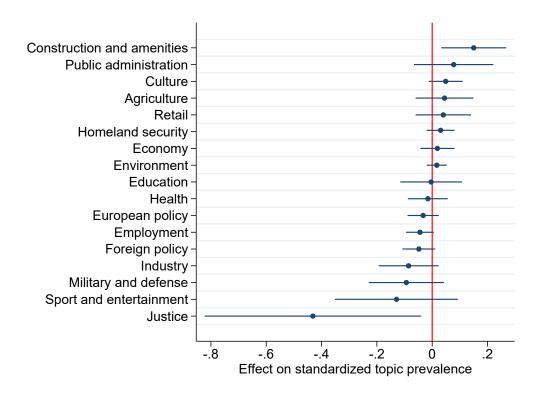
(b) Amount of corporate donations

Notes: This figure shows the coefficients and their 95% confidence intervals from a regression of the total number of corporate donations (Figure D.16a) or the total amount of corporate donations in constant euros per voter (Figure D.16b) received in the district (summed over all the candidates) on a set of candidate characteristics averaged at the district-level (not shown) and district characteristics. Non-dichotomous explanatory variables are standardized. We use one observation per district in 1993. Standard errors are robust.
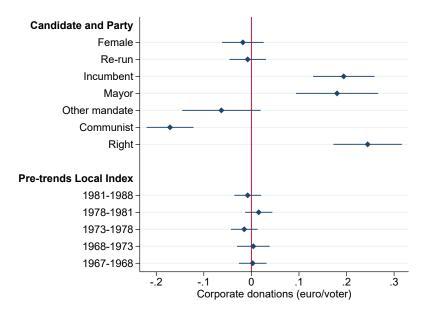
Figure D.16: District-level determinants of corporate donations in 1993, Considering the overall amount and number of corporate donations received in the district (summed over all candidates)
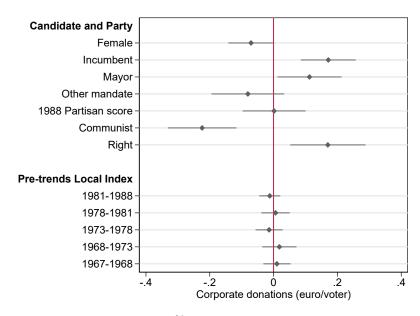
Notes: The figure shows the coefficients and their 95% confidence intervals from a regression policy topic prevalence on corporate donations. We use one observation per candidate per year. The sample includes all candidates who run both in 1993 and 1997, and whose manifesto is available. In column 6, the sample is further restricted to candidates affiliated with the five main party organizations. We control for candidate fixed effects and party×year fixed effects, as well as time-varying individual controls: indicator variables for having ran in the past, for being the incumbent, and for holding other electoral mandates. The amount of corporate donations per voter is divided by its standard deviation in 1993. The outcome is the predicted probability, for each policy topic, that a candidate manifesto focuses primarily on that topic – based on the words it contains. It is standardized by year to facilitate the comparison across topics with different levels of mean prevalence.

Figure D.17: Impact of corporate donations on policy topics in the manifestos

(a) All candidates
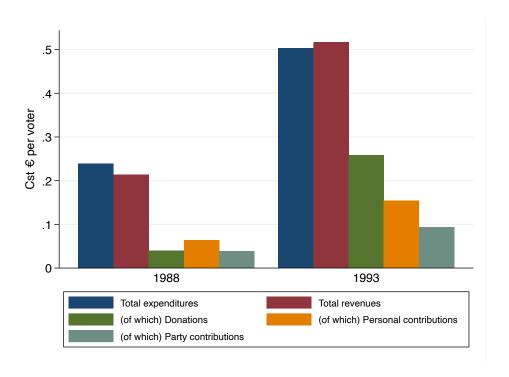
(b) Candidates running in 1988 and 1993

**Notes:** The figure shows the coefficients and their 95% confidence intervals from a regression of the amount of corporate donations in constant euros per voter received by each candidate on a set of party fixed effects, candidate characteristics and pre-trends in local index at department × party level. We use one observation per candidate in 1993. In Figure D.18a the sample includes all candidates from the Communist, the Socialist or the right-wing party (omitting candidates from the Socialist party). In Figure D.18b the sample is further restricted to candidates who run both in 1988 and 1993. Standard errors are clustered at the district level.

Figure D.18: Corporate donations and trends in local index before 1988

Notes: The figure provides summary statistics on candidates' expenses and revenues at the 1988 and 1993 elections. The data cover the sub-sample of candidates who ran both in 1988 and in 1993 at the legislative elections in the Seine Maritime department.

Figure D.19: Candidates' accounts: 1988 and 1993, Anecdotal evidence from Seine Maritime

E   Additional Tables to Chapter 2

Table E.1: Summary statistics: corporate donations in 1993, Sub-sample of candidates who received at least one corporate donation

	Mean	St.Dev	Min	Max	N
# Corp. Donations	8.79	9.49	1.00	63.00	1,701
Corp. Donations in cst euros	24,406	30,026	2	330,208	1,701
Corp. Donation (euro/voter)	0.37	0.48	0.00	6.28	1,701
% Corp. Donations in total revenue	37.44	27.73	0.00	100.00	1,701

Notes: The table presents summary statistics on corporate donations received by candidates in 1993. An observation is a candidate and the sample includes candidates who received at least one corporate donations. Other notes as in Table 2.1.

Table E.2: Summary statistics: Corporate donations in 1993, at the district level

	Mean	St.Dev.	Min	Max	N
Electoral district					
Registered voters	68,238	11,293	26,468	111,715	555
# Candidates	9	2	5	18	555
# Candidates with Corp. Donations	3	1	0	8	555
Corporate donations					
# Corp. Donations	26	19	1	109	555
Mean Corp. Donations	2,239.65	1,256.23	0	8,479	555
Total Corp. Donations	53,786.89	40,162.12	0	218,872	555
Total Corp. Donations					
Small donors	1,690.70	1,279.12	10	9,842	545
Multiple donors	3,107.58	1,648.21	0	9,842	533
Single-district donors	1,407.17	1,447.82	20	7,874	151
Multi-district donors	3,200.28	1,643.90	0	9,842	531
Left-wing parties donors	2,683.41	2,233.45	20	9,842	210
Right-wing parties donors	2,800.90	2,457.54	46	9,842	368
Non-partisan donors	3,316.84	1,889.82	197	9,842	514
Share Corp. Donations					
Small donors	0.46	0.23	0	1	555
Multiple donors	0.54	0.23	0	1	555
Single-district donors	0.02	0.06	0	1	555
Multi-district donors	0.51	0.24	0	1	555
Left-wing parties donors	0.03	0.06	0	0	555
Right-wing parties donors	0.10	0.14	0	1	555
Non-partisan donors	0.44	0.23	0	1	555

Notes: The table presents summary statistics on electoral district and corporate donations in 1993 at the district level. Mean and total corporate donations are in constant euros. Total Corp. Donations is the sum of corporate donations in a district in constant euros. Share Corporate Donations is the share of corporate donations out of total revenues in a district. Small donors are donors who made only one donation in 1993, multiple donors made more than one donation. Single-district donors (resp. multi-district) are donors who gave to candidates running in the same district (in more than one district). Left-wing parties (resp. right-wing parties) donors are multiple donors who made all the donations to candidates endorsed by left-wing (resp. right-wing) parties, non-partisan donors are multiple donors who gave to both left-wing and right-wing candidates.

Table E.3: Largest corporate donors in 1993

Donor Name	Sum donation (cst euros)	# Donations
COLAS	401367.8	96
BOUYGUES	314952.6	47
SOGEA	312590.5	82
SPIE	304126.1	59
SAUR	258851.7	62
SCREG	244875.7	60
SOCIETE DES EAUX	225781.7	53
DUMEZ	168302.8	35
CAMPENON BERNARD	165350.1	38
OMNIUM	163184.8	38
VIA TRANSPORT	139760.2	31
GTM TP	120075.7	23
SAE	119091.5	21
SODEXHO	116926.2	21
BEUGNET	113776.6	31
ESSYS MONTENAY	106296.5	25
STREICHENBERGER	101965.9	26
JEAN LEFEBVRE	92763.39	41
SUPAE	90548.88	14
MONOPRIX	87989.89	18

Notes: The table presents the largest 20 donors in 1993, the number of donations and the amount they spent in the campaign. Total donations are in constant euros.

Table E.4: Comparison of donation patterns across donors' characteristics

	Mean (a)	N (a)	Mean (b)	N (b)	Diff	p-value
# Donations						
(a) Left vs (b) right-only donors	3.05	404	2.49	1,044	0.56	0.00
(a) Left or right-only vs (b) non-partisan donors	2.62	1,562	14.74	4,110	-12.12	0.00
(a) Non-identified sector vs (b) identified sector	1.34	6,704	8.29	7,780	-6.95	0.00
Mean donation (cst euros)						
(a) Left vs (b) right-only donors	2,574.86	404	2,615.47	1,044	-40.61	0.80
(a) Left or right-only vs (b) non-partisan donors	2,584.12	1,562	3,139.83	4,109	-555.71	0.00
(a) Non-identified sector vs (b) identified sector	1,216.79	6,701	2,795.81	7,780	-1,579.03	0.00

Notes: This table compares the number of donations and the mean donations of corporate donors in 1993 included in our sample across the following criteria: if multiple donors gave to left-wing of right-wing parties only, if they gave to one side only (left or right) or both (non-partisan), and if their sector of activity was found during the cleaning procedure or not. For each observed donor characteristic, we report mean values and number of non-missing observations for each group (a) and (b), the difference in mean values between the two groups and the *p-value* associated with the test that this difference is zero.

Table E.5: Summary statistics by sector of activity

	mean	sd	min	max	count
Agriculture					
Mean donation	1,225.65	2,242.82	6	9,842	184
Sum donations	1,790.31	4,155.91	6	37,401	184
Construction					
Mean donation	2,295.53	2,584.94	20	10,138	1,615
Sum donations	6,151.99	20,623.13	20	401,368	1,615
Culture					
Mean donation	1,908.16	2,576.25	20	9,842	157
Sum donations	2,448.34	4,720.40	20	49,211	157
Economy-Finance					
Mean donation	2,454.39	3,007.26	6	9,842	586
Sum donations	3,711.16	5,499.71	6	39,369	586
Environment-Energy					
Mean donation	3,576.80	2,760.29	30	9,842	160
Sum donations	12,990.71	32,432.74	30	304,126	160
Health					
Mean donation	1,825.77	2,823.38	10	9,842	256
Sum donations	2,793.78	6,624.57	10	76,770	256
Industry					
Mean donation	2,198.35	2,797.01	10	29,527	746
Sum donations	4,402.45	10,967.34	10	163,185	746
Justice					
Mean donation	757.85	818.28	98	2,362	10
Sum donations	757.85	818.28	98	2,362	10
NGOs					
Mean donation	3,908.34	3,153.35	49	9,842	35
Sum donations	7,305.05	8,189.27	49	36,416	35
Retail					
Mean donation	1,963.44	2,681.33	10	9,842	805
Sum donations	3,150.20	7,770.07	10	116,926	805
Sport					
Mean donation	1,074.95	2,231.77	20	9,842	23
Sum donations	1,661.20	4,084.58	20	17,716	23
Travel					
Mean donation	478.76	489.81	39	1,968	21
Sum donations	576.01	583.17	39	1,968	21
Unknown					
Mean donation	1,189.47	1,623.73	10	9,842	5,870
Sum donations	1,389.04	2,088.76	10	47,243	5,870
Total					
Mean donation	1,632.80	2,234.93	6	29,527	10,468
Sum donations	2,856.81	10,277.75	6	401,368	10,468

Notes: An observation is a donor in 1993. Donations are in constant euros.

Table E.6: Left-right words

Left	Right
dividend	terrorist
antidemocratic	murderer
poverty	criminal
disarmament	foreigner
benefits	europe
thatcher	taxation
emancipation	persecution
victory	independence
law	identity
xenophobia	utopia

Notes: This table shows examples of words – translated in English – with lowest (left-wing) and highest (right-wing) ideological scores, both in 1993 and in 1997. These scores (or loadings) are obtained by fitting a multinomial inverse regression of word frequency in manifestos on an indicator variable equal to one if the candidate is from a well-identified right-wing party as opposed to a well-identified left-wing party – for 1993 and 1997 separately.

Table E.7: Topic-specific words

Homeland security	Education	Environment	Retail	Health
vote by proxy	geology	fishermen	bakery	speech therapy
police	tenure	birds	hairdresser	paramedical
firefigther	bilingual	game (animals)	craftmanship	hepatitis
electoral	school district	hunting	butcher	spokesperson
homeland	school board	fauna	slaughterhouse	physical therapy
passport	academia	waste	retail	transfusion
library	geography	farming	organic	addict
tobacco shops	teacher	pollution	tobacco shops	midwife
algerian	trainer	gas	business	surgery
violation	biology	taxi	training	anesthesy

Economy	Construction and amenities	Public administration	Employment	Justice
tobacco shop	national road	decentralisation	healthcare	seal
gas	river	rank	job training	clerk
bank customer	tourism	library	pension	prosecutor
value added	railroad	secretary	job seeking	prison
slaughterhouse	gas	assignment	disabled	lawyer
butcher	traveler	territory	solidarity	accountable
retail	freeway	city hall	trainee	magistrate
russian	aviation	citizenship	unemployment benefits	jurisdiction
deductible	car	exam	occasional worker	justice
taxation	traffic	application	internship	offense

Agriculture	Military and defense	Foreign policy	Industry	Culture
sheep	officer	execution	telecommunications	archeology
farmers	veteran	arrest	postal service	library
pig	prisonner	torture	gas provider	bicentennial
fishing	resistance	russian	textile	disc
milk	police	amnesty	electricity	french speaking
cereals	army	united nations	energy	movie theater
cow	troop	french speaking	oil	museum
vegetable	mutilation	diplomacy	diversification	culture
flock	deportation	turkey	industry	channel
harvest	defense	foreign	phone	music

Table E.7: Topic-specific words (cont.)

Sport and entertainment	European policy
olympic games	turkey
soccer	english
ski	textile
youth	parliament
sport club	translation
physical education	trade agreement
swimming pool	cereals
amateur	belgian
organizer	greek
alcohol	common agricultural policy

Notes: This table shows, for each policy topic, examples of words – translated in English – with highest topic loadings. These loadings are obtained by fitting a multinomial inverse regression of word frequency in written questions to the government on a set of dummies incating which topic (based on targeted Ministry) the questions are adressed to.

Table E.8: Prevalence of policy topics in candidate manifestos

	Mean	sd
Topic		
Agriculture	2.36	6.10
Construction and amenities	3.89	6.71
Culture	1.72	2.54
Military and defense	3.47	6.13
Economy	5.01	7.68
Education	5.29	8.72
Employment	14.31	15.74
Environment	4.17	13.23
European policy	0.57	1.61
Foreign policy	8.11	9.86
Health	3.75	4.16
Industry	1.75	2.78
Homeland security	28.48	24.49
Justice	0.32	1.47
Retail	0.20	0.79
Public administration	0.16	0.37
Sport and entertainment	0.26	0.94

Notes: The table displays the mean and standard deviation for the prevalence of each policy topic, defined as the predicted probability (in percentage points) that a candidate manifesto focuses primarily on that topic. The sample contains all first round manifestos from 1993 and 1997. N=12,673.

Table E.9: Summary statistics for covariates at the district level

	Mean	sd	Min	Max	Count
# Municipalities in the district	62.83	61.46	1	342	555
Region capital in the district	0.10	0.29	0	1	555
Urban district	0.25	0.43	0	1	555
Census 1990					
No diploma	47,264	41,845	3,521	358,972	555
Higher education	9,491	11,486	280	70,057	555
Agriculture	1,165	1,233	0	6,056	555
Blue-collar worker	11,090	7,474	604	61,394	555
65+ years old	16,320	16,467	1,052	134,100	555
25-34 years old	17,390	15,029	1,128	118,764	555
Covariates 1993					
District municipalities revenues	227,104	736,528	0	3,843,893	555
Number of firms	3	10	0	55	555
Mean number of employees per municipality	53.76	173.03	0	917	555
Total payroll (in thousand euros)	8,691.32	30,619.26	0	161,998	555
% employees in top 1%	0	2	0	8	555
Covariates 1997					
District municipalities revenues	266,059.67	871,395.55	0	4,552,347	555
Number of firms	4	11	0	61	555
Mean number of employees per municipality	54	173	0	918	555
Total payroll (in thousand euros)	9,309.73	32,369.09	9	171,363	555
% employees in top 1%	0.45	1.48	0	8	555

Notes: The table presents summary statistics on district covariates. An observation is a district. Census in 1990 are municipality-level census data averaged at the district level. Covariates in 1993 and 1997 are from the revenues and annual spending in infrastructure of the French municipalities with more than 10,000 inhabitants summed at the district level (district municipalities revenues and operating expenses) and from the "Déclaration Annuelle de Données Sociales" (DADS), a detailed French database on wages, summed at the district level (number of firms, employees per municipality, total payroll, share on employees in the top 1% of revenues. Municipalities' revenues and payroll are in constant euros.

Table E.10: Summary statistics: corporate donations in 1993, Sub-sample of candidates who run both in 1993 and 1997

	Mean	St.Dev	Min	Max	N
Corp. Donations > 0	0.46	0.50	0.00	1.00	1,425
# Corp. Donations	4.98	9.01	0.00	63.00	1,425
Corp. Donations in cst euros	14,822	26,750	0	201,274	1,425
Corp. Donation (euro/voter)	0.22	0.41	0.00	3.46	1,425
% Corp. Donations in total revenue	18.47	26.71	0.00	98.23	1,425

Notes: The table presents summary statistics on corporate donations received by candidates in 1993. An observation is a candidate and the sample includes candidates who run both in 1993 and 1997. Other notes as in Table 2.1.

Table E.11: Robust impact on different samples, depending on the availability of donations data

(a) Disaggregated donations unavailable

	Local index	Local references	National references	Left-right score	Extremeness	Originality index
	(1)	(2)	(3)	(4)	(5)	(6)
Corporate donations	0.161***	0.249***	-0.132***	-0.001	-0.005	0.033**
	(0.030)	(0.054)	(0.050)	(0.005)	(0.004)	(0.014)
Observations	2620	2620	2620	2620	2620	2088
Mean outcome after ban	-0.661	1.197	2.662	-0.017	0.741	-2.154
R2-Within	0.029	0.023	0.008	0.004	0.008	0.008

(b) Disaggregated donations equal to aggregate amount

	Local index	Local references	National references	Left-right score	Extremeness	Originality index
	(1)	(2)	(3)	(4)	(5)	(6)
Corporate donations	0.172***	0.279***	-0.144**	-0.005	-0.005	0.026
	(0.051)	(0.098)	(0.071)	(0.007)	(0.006)	(0.022)
Observations	1968	1968	1968	1968	1968	1472
Mean outcome after ban	-0.769	0.987	2.700	-0.026	0.906	-2.360
R2-Within	0.017	0.022	0.006	0.008	0.007	0.007

Notes: Standard errors are clustered by district and shown in parentheses (***, **, * indicate significance at 1, 5, and 10 percent, respectively). Panel (a) includes all candidates for whom the aggregate amount of corporate donations is available but the data on disaggregated donations is not. Panel (b) includes candidates for whom the aggregate amount of corporate donations is exactly equal to the sum of individual corporate donations from the *Journal Officiel*. Other notes as in Table 4.1.

Table E.12: Impact of corporate donations on broad policy topics by party type

	Homeland and administration	Foreign policy	Economy	Social
	(1)	(2)	(3)	(4)
Mainstream*Corp.Don.	0.719	-0.279**	1.104**	-0.904
	(0.520)	(0.118)	(0.537)	(0.574)
Niche*Corp.Don.	-9.414	-1.600*	39.819***	-12.702**
	(12.160)	(0.877)	(8.666)	(5.968)
Independent*Corp.Don.	-2.284	-0.239	6.993***	-4.250
	(2.397)	(0.842)	(2.080)	(2.788)
Observations	2602	2602	2602	2602
Mean outcome after ban	15.189	3.011	28.788	34.827
R2-Within	0.008	0.004	0.032	0.010

Notes: The outcome is the predicted probability, for each policy topic, that a candidate manifesto focuses primarily on that topic out of 4 broad topics – based on the words it contains. It is measured in percentage points. Mainstream parties are the Communist, Socialist and right-wing parties. Niche parties are the Green and far-right parties as well as smaller parties. Independent candidates are not affiliated with any national party. Other notes as in Tables 4.1 and 4.3.

Table E.13: Impact of corporate donations on campaign communication by candidate type

	Local index	Local references	National references	Left-right score	Extremeness	Originality index
	(1)	(2)	(3)	(4)	(5)	(6)
Corporate donations	0.185***	0.342***	-0.113	0.001	-0.015*	0.020
	(0.058)	(0.110)	(0.087)	(0.009)	(0.008)	(0.024)
Corp.Don.*Female	-0.064	-0.033	0.133	-0.016	0.008	-0.000
	(0.088)	(0.144)	(0.164)	(0.011)	(0.011)	(0.030)
Corp.Don.*Re-run	-0.146	-0.259	0.134	0.007	0.026*	-0.013
	(0.095)	(0.184)	(0.145)	(0.015)	(0.013)	(0.036)
Corp.Don.*Incumbent	0.100	0.087	-0.172	-0.005	-0.016	0.038
	(0.084)	(0.150)	(0.142)	(0.013)	(0.012)	(0.030)
Corp.Don.*Mayor	0.083	0.146	-0.061	-0.006	0.006	0.001
	(0.061)	(0.108)	(0.107)	(0.011)	(0.012)	(0.027)
Corp.Don.*Other mandates	0.127	0.227**	-0.063	-0.027**	-0.001	-0.037
	(0.077)	(0.116)	(0.193)	(0.013)	(0.012)	(0.038)
Observations	2602	2602	2602	2602	2602	2070
Mean outcome after ban	-0.658	1.199	2.658	-0.021	0.740	-2.152
R2-Within	0.036	0.031	0.010	0.006	0.010	0.009

Notes: The amount of corporate donations per voter (divided by its standard deviation in 1993) is interacted with indicator variables for being a woman, for having ran in the past, for being the incumbent, for being a mayor and for holding any other electoral mandate (senator, departmental mandate or European MP) in 1993. Other notes as in Table 4.1.

Table E.14: Heterogeneous effect on local index by donor size

	Local index		
	(1)	(2)	(3)
Corp.Don from: small donors ≤2	0.070*		
	(0.036)		
Corp.Don from: multiple donors >2	0.061**		
	(0.027)		
Corp.Don from: small donors ≤3		0.070*	
		(0.036)	
Corp.Don from: multiple donors >3		0.062**	
		(0.026)	
Corp.Don from: small donors ≤5			0.090**
			(0.036)
Corp.Don from: multiple donors >5			0.040
			(0.026)
Observations	2602	2602	2602
Mean outcome after ban	-0.658	-0.658	-0.658
R2-Within	0.022	0.022	0.022

Notes: We define small donors as donors who make 1 or 2 donations (column 1), up to 3 donations (column 2), and up to 5 donations (column 3). Other notes as in Table 5.1.

Table E.15: Heterogeneous effect on frequency of local references by sources of funding and type of donor

	Frequency of local references			
	(1)	(2)	(3)	(4)
Corporate donations	0.283***			
	(0.055)			
Individual donations	-0.027			
	(0.053)			
Personnal contributions	0.041			
	(0.033)			
Party contributions	0.089			
	(0.065)			
Donations from small donors		0.137**	0.132*	0.134*
		(0.069)	(0.068)	(0.069)
Donations from multiple donors		0.127**		
		(0.054)		
Multiple donors: multi-districts			0.102*	
			(0.053)	
Multiple donors: single-district			0.155**	
			(0.068)	
Multiple donors: left-only				-0.061
				(0.071)
Multiple donors: right-only				0.080
				(0.056)
Multiple donors: non-partisan				0.110**
				(0.049)
Observations	2602	2602	2602	2602
Mean outcome after ban	1.199	1.199	1.199	1.199
R2-Within	0.027	0.024	0.029	0.028

Notes: The outcome is the normalized frequency of local references measured in percentage points. Other notes as in Table 5.1.

Table E.16: Heterogeneous effect on frequency of national references by sources of funding and type of donor

	Frequency of national references			
	(1)	(2)	(3)	(4)
Corporate donations	-0.148**			
	(0.058)			
Individual donations	-0.029			
	(0.057)			
Personnal contributions	-0.019			
	(0.039)			
Party contributions	-0.024			
	(0.056)			
Donations from small donors		-0.061	-0.060	-0.069
		(0.055)	(0.055)	(0.055)
Donations from multiple donors		0.004		
		(0.056)		
Multiple donors: multi-districts			0.010	
			(0.055)	
Multiple donors: single-district			-0.035	
			(0.042)	
Multiple donors: left-only				-0.028
				(0.030)
Multiple donors: right-only				0.082*
				(0.048)
Multiple donors: non-partisan				-0.032
				(0.053)
Observations	2602	2602	2602	2602
Mean outcome after ban	2.658	2.658	2.658	2.658
R2-Within	0.008	0.004	0.005	0.007

Notes: The outcome is the normalized frequency of national references measured in percentage points. Other notes as in Table 5.1.

Table E.17: Heterogeneity by donor's sector of activity

	Local index	Local references	National references
	(1)	(2)	(3)
Corp.Don from: other sectors	0.006	0.061	0.044
	(0.029)	(0.054)	(0.041)
Corp.Don from: construction	0.006	-0.007	-0.014
	(0.029)	(0.054)	(0.052)
Corp.Don from: economy	-0.001	0.022	0.020
	(0.033)	(0.058)	(0.048)
Corp.Don from: environment	0.051*	0.103**	-0.024
	(0.030)	(0.050)	(0.052)
Corp.Don from: industry	0.008	0.007	-0.020
	(0.030)	(0.053)	(0.051)
Corp.Don from: retail	-0.013	0.056	0.094**
	(0.031)	(0.058)	(0.046)
Corp.Don from: unknown	0.103**	0.134	-0.117*
	(0.045)	(0.085)	(0.061)
Observations	2602	2602	2602
Mean outcome	-0.658	1.199	2.658
R2-Within	0.027	0.029	0.011

Notes: The amount of corporate donations per voter received by each candidate is broken down into amounts received by donors form different sectors of activity. Other notes as in Table 4.1, columns 1-3.

Table E.18: Impact of corporate donations on local prevalence, Sub-sample of elected representatives

	Local index	Local references	National references
	(1)	(2)	(3)
Corporate donations	0.113**	0.186**	-0.071
	(0.045)	(0.078)	(0.084)
Observations	448	448	448
Mean outcome after ban	-0.361	1.804	2.605
R2-Within	0.040	0.032	0.011

Notes: The sample is restricted to politicians elected both in 1993 and 1997. Other notes as in Table 4.1.

Table E.19: Impact of corporate donations on interventions in low- and high-visibility debates

(a) Low-visibility debates

	Number of interventions	Local index	Local references	National references
	(1)	(2)	(3)	(4)
Corporate donations	0.414	0.091	-0.043	-0.303
	(0.627)	(0.089)	(0.032)	(0.246)
Observations	222	214	214	214
Mean outcome	6.617	-1.332	0.254	3.096
R2-Within	0.088	0.034	0.053	0.032

(b) High-visibility debates

	Number of interventions	Local index	Local references	National references
	(1)	(2)	(3)	(4)
Corporate donations	2.129	-0.110	-0.016	0.314
	(3.116)	(0.076)	(0.025)	(0.251)
Observations	330	322	322	322
Mean outcome	38.233	-1.428	0.226	3.771
R2-Within	0.050	0.047	0.004	0.045

Notes: We distinguish interventions made in low-visibility debates (generating a below-median number of interventions) from interventions made in high-visibility debates (generating an above-median number of interventions). Other notes as in Table 5.2, Panel b.

Table E.20: Impact of corporate donations on broad policy topics in legislative discourse

(a) Written questions to the government

	Homeland and administration	Foreign policy	Economy	Social
	(1)	(2)	(3)	(4)
Corporate donations	0.123	0.002	-0.568	0.335
	(0.165)	(0.002)	(0.527)	(0.481)
Observations	416	416	416	416
Mean outcome	8.828	0.082	60.105	27.456
R2-Within	0.059	0.016	0.033	0.012

(b) Debate interventions

	Homeland and administration	Foreign policy	Economy	Social
	(1)	(2)	(3)	(4)
Corporate donations	-0.906	0.521	-1.962	1.703
	(0.904)	(0.871)	(1.424)	(1.590)
Observations	356	356	356	356
Mean outcome	14.107	8.962	31.993	28.224
R2-Within	0.030	0.042	0.027	0.022

Notes: Same notes as in Tables 5.2 and E.12.

Table E.21: Robust impact of corporate donations on the local index

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Corp.Don.	0.162***				0.272***	0.159***	0.162***	0.137***	0.096***	
	(0.033)				(0.056)	(0.040)	(0.030)	(0.030)	(0.022)	
Log Corp.Don.		0.183***								
		(0.031)								
Receiving any Corp.Don.			0.238***							
			(0.080)							
Number of Corp.Don.				0.022***						
				(0.005)						
Corp.Don.^{2}					-0.021***					
					(0.008)					
Share Corp.Don./Revenue										0.008***
										(0.002)
Observations	2602	2602	2602	2602	2602	2602	2602	2602	5430	2518
Mean outcome after ban	-0.658	-0.658	-0.658	-0.658	-0.658	-0.658	-0.658	-0.658	-0.730	-0.660
R2-Within	0.030	0.032	0.014	0.028	0.032	0.036	0.055	0.044	0.013	0.028
Candidate FE	✓	✓	✓	✓	✓	✓	✓	✓		✓
Party*Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
District*Year FE						✓				
Party*District FE									✓	
Main controls	✓	✓	✓	✓	✓	✓	✓		✓	✓
District controls							✓			
Controls*Year FE								✓		
Larger clusters	✓									

Notes: Standard errors are clustered by department (column 1) or by district (columns 2 through 10) and shown in parentheses (***, **, * indicate significance at 1, 5, and 10 percent, respectively). We use one observation per candidate per year. In all columns but column 9, the sample includes all candidates who run both in 1993 and 1997, and whose manifesto is available. In column 9, the sample includes all candidates who run for a party that was present in the same district both in 1993 and 1997, and excludes independent candidates without a clear party affiliation. In all columns, the outcome is the local index of each candidate manifesto, which measures the prevalence of local references over national ones (divided by its standard deviation). In columns 1-2 and 5-9, the amount of corporate donations per voter and the log of this amount (plus one) are divided by their respective standard deviation in 1993. In column 10, the share of corporate donations in a candidate's total revenue is measured in percentage points. We control for candidate fixed effects and party×year fixed effects, as well as individual controls: indicator variables for having ran in the past, for being the incumbent, and for holding other electoral mandates. In column 6, we add district×year fixed effects. In column 7, we control for time-varying district characteristics. In column 8, the main candidate controls are interacted with the year fixed effects, as well as past controls measured in 1988. In column 9, candidate fixed effects are replaced with party×district fixed effects.

Table E.22: Impact of corporate donations on local prevalence, Sub-sample of candidates present in 1988

	Local index	Local references	National references	Left-right score	Extremeness	Originality index
	(1)	(2)	(3)	(4)	(5)	(6)
Corporate donations	0.138***	0.203***	-0.114*	-0.001	-0.004	0.039**
	(0.037)	(0.063)	(0.067)	(0.007)	(0.006)	(0.016)
Observations	1106	1106	1106	1106	1106	1074
Mean outcome after ban	-0.543	1.528	2.717	0.096	0.563	-1.985
R2-Within	0.052	0.024	0.031	0.003	0.014	0.014

Notes: The sample is restricted to candidates who run in 1988, 1993 and 1997. Other notes as in Table 4.1.

Table E.23: Comparison of included and excluded observations

	Mean included	N included	Mean excluded	N excluded	Diff	p-value
Female	0.14	1,414	0.22	3,668	-0.08	0.00
Re-run	0.41	1,414	0.15	3,668	0.26	0.00
Incumbent	0.19	1,414	0.04	3,668	0.15	0.00
Mayor	0.07	1,414	0.02	3,668	0.05	0.00
Other mandates	0.04	1,414	0.02	3,668	0.02	0.00
Revenues (euro/voter)	0.54	1,414	0.27	3,668	0.28	0.00
Corp.Don. (euro/voter)	0.22	1,414	0.08	3,668	0.14	0.00
Indiv.Don. (euro/voter)	0.06	1,414	0.03	3,668	0.03	0.00
Personnal.contrib. (euro/voter)	0.09	1,414	0.07	3,668	0.02	0.00
Party.contrib (euro/voter)	0.14	1,414	0.07	3,668	0.07	0.00

Notes: The table compares candidates included in our sample (i.e. candidates who ran both 1993 and 1997) to excluded ones. For each observed candidate characteristic and source of campaign revenue, we report mean values and number of non-missing observations for each group, the difference in mean values between the two groups and the *p-value* associated with the test that this difference is zero.

Table E.24: Impact of corporate donations on selection into sample

	Re-runner	Manifesto available
	(1)	(2)
Corporate donations	0.034***	-0.001
	(0.010)	(0.005)
Observations	11308	2828
Mean outcome after ban	0.199	0.929
R2-Within	0.003	0.012
District FE	✓	
Candidate FE		✓
Controls		✓

Notes: Standard errors are clustered by district and shown in parentheses (***, **, * indicate significance at 1, 5, and 10 percent, respectively). We use one observation per candidate per year. In column 1 the outcome is an indicator variable indicating if the candidate ran again in the next election (in the same district and for the same party). We control for district fixed effects and party×year fixed effects, as well as individual controls: indicator variables for being a woman, having ran in the past, for being the incumbent, and for holding other electoral mandates. In column 2 the outcome is an indicator variable indicating if the candidate has a first-round manifesto available and the sample includes candidates who ran both in 1993 and 1997. We control for candidate fixed effects and party×year fixed effects as well as time-varying individual controls.

Table E.25: Impact of corporate donations on total revenue and other sources of revenue

	Total revenue	Donations from individuals	Party contributions	Personal contributions
	(1)	(2)	(3)	(4)
Corp.Don. (euro/voter)	0.735***	-0.054***	-0.138***	-0.112***
	(0.026)	(0.016)	(0.021)	(0.020)
Observations	2828	2828	2828	2828
Mean outcome after ban	0.432	0.054	0.075	0.287
R2-Within	0.593	0.041	0.065	0.056

Notes: Standard errors are clustered by district and shown in parentheses (***, **, * indicate significance at 1, 5, and 10 percent, respectively). We use one observation per candidate and per year. The sample includes all candidates who run both in 1993 and 1997, and for whom total revenues (column 1) or different sources of revenue (columns 2-4) are known. We control for candidate fixed effects and party×year fixed effects, as well as individual controls: indicator variables for having ran in the past, for being the incumbent, and for holding other electoral mandates. The amount of corporate donations as well as all outcomes are measured in constant euros per voter.

Table E.26: Impact of corporate donations on shares of different sources in total revenue

	Donations from individuals	Party contributions	Personal contributions
	(1)	(2)	(3)
Share of corporate donations	-0.177***	-0.486***	-0.317***
	(0.021)	(0.036)	(0.038)
Observations	2726	2726	2726
Mean outcome after ban	10.829	19.375	66.783
R2-Within	0.043	0.124	0.056

Notes: The share of campaign revenue coming from each source is measured in percentage points. Other notes as in Table E.25.

Chapter 3

TV in times of Political Crisis: Evidence from the 2017 Presidential Election in Kenya

Abstract

What is the impact of television in periods of political crisis? This paper studies the nullification of the 2017 presidential election in Kenya and the organization of a repeat ballot, which triggered a wave of protests and created a situation of grave uncertainty. I exploit the expansion of digital television between 2013 and 2017 and use the distance to the transmitters and terrain features to exhibit exogenous variation in signal coverage. I then estimate a difference-in-differences model using as outcomes turnout in each of the two separate ballots. I find that TV did not impact turnout in the first election but that it heterogeneously influenced voters in the repeat election. Turnout decreased by 4% in pro-opposition regions but was 7% higher in pro-government bastions. Using text analysis methods, I document Kenyan TV channels' coverage of the crisis, and explore the interaction between *direct* and *indirect* exposure to conflict. I argue that social coordination through collective exposure to TV could amplify the influence of the broadcasts.

1 Introduction

How do the media affect voters' behavior in times of political uncertainty? Access to information plays a key role in shaping individual preferences and can directly influence political outcomes (Gentzkow and Shapiro, 2006; Durante and Knight, 2012; Gavazza et al., 2019). In periods of political uncertainty or conflict, the media can contribute to the ongoing turmoil by influencing citizens' perceptions and behavior (Yanagizawa-Drott, 2014; Armand et al., 2020).

In this paper, I compare two elections that took place in a three-month window to investigate how political context influences voters' choices. More precisely, I focus on the electoral crisis that occurred in Kenya in 2017, which pitted the incumbent President against the leader of the opposition in the race for the presidential mandate. The presidential election took place in August 2017 and resulted in a victory for the incumbent. However, the Supreme Court decided to nullify the election result and ordered the electoral commission to schedule a repeat ballot.[1] The ruling marked a breakthrough in Kenyan politics, and the debate around the nullification and the organization of new polls was accompanied by with a wave of opposition-led protests conducted. While the first election was organized in a normal electoral context, the legal dispute and the riots dramatically fed a climate of crisis throughout the country. The level of political uncertainty, measured for example, by the World Uncertainty Index, rose sharply raised before the repeat election and reached its highest level for ten years.

This paper uses this unique setting to investigate the effect of exposure to TV on voters, looking at the two rounds separately. Using a difference-in-differences model, I look at polling stations that gained access to a TV signal between the last two presidential elections of 2013 and 2017, and compare the August and October ballots with the 2013 polls. Previous studies have highlighted the role of access to information in unstable political environments (Della Vigna et al., 2014; Casey, 2015; Knight and Tribin, 2019); this paper contributes to the literature by providing new evidence on how voters react to messages when political odds evolve unexpectedly.

To analyze the impact of TV on voters, I use a novel dataset on the staggered introduction of digital television in Kenya between 2013 and 2018. According to the *Communication Authority of Kenya*, more than 2,500 sublocations benefited from the migration from analog to digital television: this corresponds to approximately 7,000 polling stations that benefited

---

[1]The ruling of the Kenyan Supreme Court to nullify the election was announced on September 1st 2017 after the incumbent President had been declared the winner on August 2017. The cancellation followed a complaint lodged by the opposition party and was based on the misalignment in the reporting of the results to the electoral commission between the different polling centers.

from the expansion of the signal over the period. One challenge involved in estimating the causal impact of TV on mobilization comes from the risk of endogeneity due to plausible non-randomness in signal expansion. TV transmitters could have been installed in strategic locations - to maximize the number of receivers for instance -, and the voters with access to the signal might show specific political attitudes.

To overcome this risk of bias, I first investigate the parameters that explain whether a polling station is covered by the signal in 2017, and exhibit exogenous variation in the likelihood of receiving the signal. I construct a geo-coded dataset of Kenyan polling centers and TV transmitters and compute the fourth polynomial of the distance between each voting center and the nearest transmitter. The dataset also includes terrain ruggedness measures, distances to the capital or urban centers, urbanization status as well as demographic and economic covariates, measured at a fine-grained level. In line with the existing literature (Olken, 2009; Yanagizawa-Drott, 2014; Miner, 2015), I find that the distance to the transmitter and terrain features are the main predictors of signal reception. For instance, an increase of 1 km increase in the distance to the transmitter raises the probability of receiving the signal by around 60%. To capture the causal effect of exposure to TV, I rely on the identification assumption that locations with or without access to TV in 2017 are similar in unobserved characteristics, conditional on observables. I run a series of placebo tests to ascertain exogeneity in signal reception controlling for the determinants of signal propagation. I find that conditional on signal predictors, political preferences before 2013, and ethnic belongings, are not significantly correlated with TV reception in 2017. I run similar checks on economic and social covariates – before and at the time of the election – and do not find evidence of bias in the allocation of TV, conditional on the propagation determinants.

I then estimate a difference-in-differences model to measure the impact of access to TV on turnout for both the first and the recall election, where the treated polling stations are those that were given access to the signal between 2013 and 2017, controlling for the signal predictors previously identified. Turnout dropped from 79% to 41% because of the decrease in the remobilization rate among the opposition's supporters. First, I find no evidence that exposure to the signal changed voters' likelihood of turning out in the first election – i.e. in normal times. However, it had a sizable impact on the repeat election. Based on the stark political divides across the territory, which largely interact with ethnic belongings (Horowitz and Long, 2016), I split the sample between polling centers that voted in favor of the opposition or the incumbent President in 2013. Strikingly, the sign of the estimates is opposite in the two political camps. In pro-opposition polling stations, moving from an area without TV to an area with TV lowers participation by 4 percentage points. Conversely, I find that TV has a positive and large effect on remobilization, +7 percentage points, among

supporters of the incumbent and final winner of the election. Moreover, as the incumbent's score in the first poll increases, the effect of signal reception grows larger. The estimates suggest that a 10-percent increase in the incumbent's score adds 1.5 percentage points to the baseline coefficient.

These findings are corroborated with a series of robustness checks. The results are similar when the sample is restricted to areas that got access to the signal after 2016, which suggests that they are not driven by more urbanized places. This indicates that the treatment effect is not contaminated by the larger penetration of the Internet and social media in bigger cities or the higher likelihood to be exposed to urban riots.[2] Moreover, a placebo test confirms that the results are not spurious as I find no effect of being granted access to the signal in 2018, i.e. just after the 2017 election on mobilization.

In the last part of the paper, I consider the possible transmission channels and explore different mechanisms that could explain the gap between the August and October rounds, as well as the opposite direction of the effect in the two camps. I first study the content of the TV broadcasts to document the type of news stories that were on air at the time of the election. I construct a dataset of news items broadcast by the five main TV channels in 2017 and implement text analysis methods to measure any changes in news coverage. I provide evidence of those TV channels' strong emphasis on the climate of tension and on the ongoing riots in the country in the aftermath of the nullification of the first election. For instance, in the two weeks before the repeat election, the share of conflict-related news on TV increased by 10%. Moreover, I bring to light a drop in the diversity of reported news stories, which indicates a shift away from the usual coverage choices, and a strong focus on the political crisis.

Since the protests were largely covered on TV, I examine the possible impact that *direct* exposure to conflict could have had on voters in addition to the effect of TV exposure. I also look at the interaction of both *direct* and *indirect* exposure to conflict, i.e. when the polling station is both close to a protest hotspot and covered by the signal. Notably, the effect of the distance to the protests is small in magnitude and opposite in the camps. Pro-opposition polling centers located further away from a protest will tend to go to vote more, while turnout is decreasing in the distance to the nearest riot among the supporters of the incumbent. However, the effect of *direct* exposure to conflict is similarly offset by access to TV in the two camps. These findings add to the existing literature as they suggest that the effect of exposure to TV broadcasts on voters' choices is greater than the effect of living close to an epicenter of the crisis.

A possible explanation for this strong effect of media exposure could be the collective

---

[2]Internet penetration in polling centers without access to TV in 2016 was below 25% in 2017.

dimension in TV watching. Indeed, in Kenya, there is limited ownership of TV sets, but people are used to being regularly exposed to news broadcasts. Indeed, screens are often installed in public places such as cafes or restaurants where people meet and watch the news. Combining data from household surveys and signal availability, I uncover that individuals are more likely to frequent restaurants or cafes in areas covered by the signal, conditional on household characteristics and regional fixed-effects. This finding is consistent with a narrative of exposure to TV broadcasts as a social experience, which could contribute in the consolidation of social coordination mechanisms. The supporters of the incumbent may have been more willing to express their support for the leader collectively, as a reaction to the loud and sometimes violent contestation led by the opposition. Conversely, in the opponent's camp, where most of the protests originated, risk aversion and defiance towards the electoral system could have been fueled by exposure to the news in public places. Considered together, these results shed new light on the important role that TV can play in periods of political uncertainty.

Literature Review

This paper speaks to three main strands of the literature. First, it adds to existing studies that focus on media influence on individual and collective behavior. In a detailed review of the literature, DellaVigna and La Ferrara (2015) present the various channels that can be at play when an audience is exposed to news stories. The link between access to information and voting have been studied extensively (Gerber and Green, 2000; Lassen, 2005; Pons, 2018), notably in developing countries (Banerjee et al., 2011; Fujiwara and Wantchekon, 2013; Casey, 2015; Marx et al., 2020; León, 2017; Chong et al., 2019).[3] Several empirical works document the effects of exposure to traditional media on turnout: while some studies find a positive relationship – e.g. Stromberg (2004) and Gentzkow et al. (2011) –, Cagé and Godechot (2017) highlights that quality of information matters and that competition between outlets can lead to a decrease in participation. A depressing effect of access to information is found by Miner (2015) in the case of radio, Falck et al. (2014) and Gavazza et al. (2019) for online sources. In the case of TV, Gentzkow and Shapiro (2006); Ellingsen and Hernæs (2018) document a negative effect on turnout.[4] This paper shows evidence of the heterogeneity of the effect, and sheds light on the role of ex ante political preferences in explaining the media's influence on voters.

---

[3]While this paper focuses on news contents, entertainment and *edutainment* programs also affect individual behavior and attitudes, as shown for instance by La Ferrara (2016) or Jensen and Oster (2009).

[4]Also related is Durante et al. (2019) who show that entertainment TV also affects political preferences in Italy or Mastrorocco and Minale (2018) who use the introduction of digital TV in Italy to examine the effects of the media on perceptions of crime.

Second, this paper relates to the literature on community or ethnic belongings and political outcomes. Numerous works have stressed how outcomes can be shaped by ethnic divides (Horowitz, 1995; Easterly and Levine, 1997; Michalopoulos and Papaioannou, 2013; Miguel, 2004; Burgess et al., 2015; Alesina et al., 2016). When it comes to strategic voting, group belongings can act as a cognitive bias and influence political preferences, especially when one's co-ethnic is running (Long and Gibson, 2015). As shown by Horowitz and Long (2016) in their study of the Kenyan 2007 presidential election, uneven distribution of communities across the territory feeds a confirmation bias mechanism. Kasara (2013) also shows that in the case of the 2007 crisis, low levels of inter-ethnic trust exacerbated the conflict. In this paper, I also discuss the possible mechanisms of voter exit or voice strategies (Chong et al., 2015), thereby contributing to a wider body of empirical studies on media and conflict (Collier and Vicente, 2014; Yanagizawa-Drott, 2014; Della Vigna et al., 2014; Armand et al., 2020). Past exposure to violence has been shown to affect community involvement, civic behavior, trust (Blattman, 2009; Bellows and Miguel, 2009; Bauer et al., 2016), and risk aversion (Voors et al., 2012; Callen et al., 2014; Moya, 2018; Brown et al., 2019).[5] This paper adds to these previous findings by documenting the interaction between *direct* and *indirect* exposure to conflict.

Last of all, this paper contributes to the literature that examines the interactions between media content and political behavior. A large body of empirical works focuses on the impact of quality and quantity of information on public policies and spending (Besley and Burgess, 2002; Larcinese et al., 2011), or on accountability and corruption (Ferraz and Finan, 2008; McMillan and Zoido, 2004).[6] In particular, private and state-owned media may have diverging editorial strategies, so that viewers' exposure to different sources can translate into altered in political behavior or attitudes (Enikolopov et al., 2011; Knight and Tribin, 2019). Using content analysis methods (Gentzkow et al., 2019), I document the emphasis placed on the crisis by Kenyan TV channels and highlight the change in news diversity on TV during the period of electoral turmoil. Moreover, in sub-Saharan African countries, individual TV ownership is often low and screens are often watched in public places. This paper provides evidence of increased frequenting of public places in regions covered by a television signal and argues that the collective dimension of news watching can amplify the influence of voters through coordination mechanisms.

---

[5]In the case of the 2007 post-electoral crisis in Kenya, Dupas and Robinson (2012) show political conflict and a rise in uncertainty can translate into negative shocks on individual livelihood or income.

[6]Moreover, previous works have shown that revenues from sales and advertising (Reuter and Zitzewitz, 2006; Di Tella and Franceschelli, 2011; Angelucci and Cagé, 2019), the structure of ownership and the level of independence of media groups (Enikolopov et al., 2011; Durante and Knight, 2012) are important determinants of editorial choices.

The rest of the paper is organized as follows. Section 2 provides background information on the 2017 election in Kenya and presents the data used in the analysis. Section 3 details the empirical strategy and the main findings. I explore the possible underlying mechanisms in Section 4. Section 5 concludes.

2   Background and Data

The presidential election of 2017 marked a breakthrough in recent political history in Kenya. In this section, I first introduce the context and data I use to study electoral outcomes at the polling center level (subsection 2.1). I then present the media landscape in Kenya and describe the data on TV coverage and broadcast contents used in the analysis (subsection 2.2). I complement the dataset with a set of terrain features and administrative or socio-economic controls (subsection 2.3).

2.1   Presidential elections in Kenya

The political landscape   In Kenya, multi-party politics was re-introduced in 1991 with the repeal of section 2A of the Constitution, under which opposition parties had been banned (Branch, 2011). Since the first election in 1992, which marked the beginning of a new political era in the country, multiple candidates have been able to for a five-year presidential term.[7]   The 2017 polls are the second general election after the constitutional reform of 2010, and the election of Uhuru Kenyatta – leader of the Jubilee Alliance – in 2013. General elections in Kenya are a multi-layered political event: voters cast a ballot to elect the president of the country, but they also choose their local representatives – namely, members of parliament (MPs), senators, county governors, members of county assembly (MCAs) and women's representatives.

August 2017 General election   On August, 8, 2017, more than 19 million Kenyans were called to the polls to cast six ballots and elect the country's new representatives. In the presidential race, the Jubilee Alliance, represented by the incumbent President, Uhuru Kenyatta and his vice-president William Ruto,stood against the National Super Alliance

---

[7]Besides the implementation of a new Constitution in 2010, the presidential mandate has remained unchanged since then: the constitutional reform added new political layers with the creation of 47 counties and county assemblies, so that governors and county assembly members are now elected at the same time as the president. The constitutional change was implemented in the aftermath of the post-electoral violence that shook the country after the 2007 election (Gibson and Long, 2009; Kramon and Posner, 2011).

(NASA) led by Raila Odinga and his co-runner Kalonzo Musyoka.[8] During the months preceding the election, the atmosphere was strained due to the repeated complaints launched by the opposition against the Independent Electoral and Boundary Commission (IEBC) in charge of organizing the polls.[9]

On August, 10, 2017, the incumbent President Kenyatta was declared the winner with 54.17% of the votes. Turnout was high, reaching 79%. Notably, the win margin of one candidate over the other was very large in their respective political bastions; Appendix Figure B.1 maps the depth of this geographic clustering on the territory. However, the opposition challenged the results shortly after they were announced, primarily due to the lack of transparency in the transmission of the results, and filed a petition calling on the Supreme Court to adjudicate.

The "Supreme Bombshell" and the repeat election On September, 1, 2017, the presidential election was declared "invalid, null and void" by the Supreme Court, which ordered that new polls must be organized within 60 days. The "Supreme Bombshell", as this episode was referred to in the media, deeply shook the country: protests led by opposition leaders crystalized around Odinga's call for the resignation of IEBC officials. The opposition blamed the electoral commission for immediately validating the results and accused its members of showing allegiance to the powers already in place. Moreover, the 60-day bound was judged too narrow to ensure a credible new ballot. Tension reached a climax when the opposition called for a boycott of the repeat election.[10] The number of riots and clashes with the police increased as the day of the repeat election grew closer. In the month before the recall election, 33 people died in fights with the police. For his part, President Kenyatta, besides publicly expressing his disagreement with the Court's decision, started a new campaign to mobilize his electorate. The media widely covered this political episode, and TV news broadcasts reinforced this climate of tension and uncertainty.[11] The repeat election was held on October, 26, 2017. Turnout dropped to 36%.[12] Figure 2.1 plots turnout in the first and recall election against the win margin of the incumbent in the first election. While turnout is nearly linear across the political camps in August, participation in the recall election moves positively with the previous score of the incumbent. Because the boycott

---

[8]Six other candidates run either for small or independent parties, together winning less than 1% of the total vote share.

[9]One bone of contention was the electoral technology for voter identification and reporting of the results, which was perceived as being open to manipulation by the opposition (Wairuri, 2017).

[10]Appendix Figure B.2 shows two headlines from Kenyan newspapers to illustrate the climate of high tension after the Court's statement.

[11]Subsection 4.1 provides a detailed analysis of news broadcast over the period.

[12]Excluding the polling centers that were blocked on the day of the repeat election, turnout was 41%.
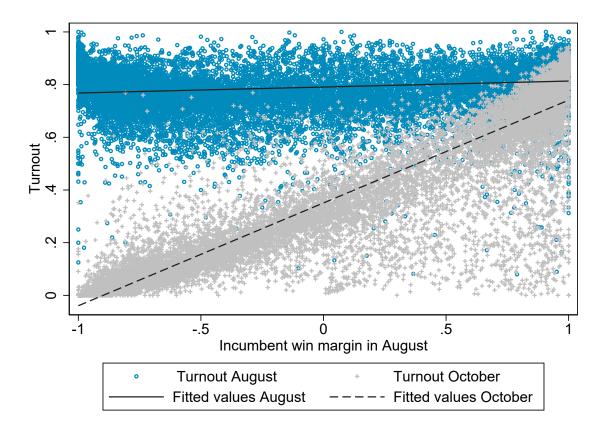
was widely followed by the opposition supporters, the second vote effectively turned into a one-candidate election: the incumbent President Kenyatta was proclaimed winner by the Supreme Court with 98% of the votes cast.[13]

The Supreme Court's decision was unprecedented and constituted a major event in recent Kenyan politics. Appendix Figure B.3 shows that political uncertainty – as measured by the World Uncertainty Index, often used in the policy or finance literature (see for instance Baker et al., 2016) – was at its highest level for a decade during the quarter in which the repeat election wad held. I use this unique setting of a sudden change in the political odds to compare mobilization patterns in normal times, i.e. the August election of August, and in times of crisis, the October recall election. Notably, due to the boycott called for by the opposition leaders, the outcome of the repeat election was almost a foregone conclusion – which partly explains the drop in turnout. Despite this, President Kenyatta succeeded in mobilizing a substantial share of his electorate. Moreover, the variation in voter participation in pro-government areas raises the question of the determinants of collective mobilization to signal one's support for a political leader.

Electoral data To study the results of the 2017 election at the polling station level, I collected the scans of the result forms – known as forms 34B –, issued by the IEBC. I digitized the scanned copies of these forms which are filled in by the constituency electoral officers, who report the number of voters, votes cast and votes rejected for each polling station. Appendix Figure B.4 shows a sample of form 34B. For both the first and the repeat election, I used the unique identifier code attributed to polling stations to merge the data with the electoral commission's official records. To the best of my knowledge, this dataset is the first to reconcile the two rounds of the 2017 Kenyan presidential election at the polling station level.

A polling center can be split into several polling stations: there are 40,770 stations for 24,062 centers (1.7 stations per center on average). With the impact of TV signal reception on turnout as the main research question, I aggregate the results at the polling center level. Finally, I merge this dataset with the GIS coordinates of the centers. While the quality of the scans was good enough for the majority of forms 34B, it was not possible to digitize some of them because of missing information or too low resolution (e.g. incomplete polling station codes, missing or unreadable result columns, etc). This triggers slight attrition in the final dataset (13%), which consists of a total of 35,889 polling stations out of the 40,488 officially

---

[13]Several polling stations in the Western counties of Kisumu, Migori, Siaya and Homa Bay, remained closed on election day. Several outbursts of violence were observed between the two rounds, adding to the preexisting higher level of tensions in the year before the election: 244 episodes of violence were recorded between the two rounds, and 870 in the year preceding the October election.

Notes: The figure plots turnout in August (blue circles) and in October (grey crosses) against the win margin of the incumbent President Kenyatta (Jubilee Alliance) in the first election of August 2017. An observation is a polling station.

Figure 2.1: Turnout in August and in October 2017

recorded by the IEBC, pertaining to 21,021 centers. Table 2.1 shows summary statistics on these electoral outcomes.[14]

To compare the 2017 election with the last presidential ballot held in 2013 and estimate a difference-in-differences model, I merge the abovementioned digitized data with a dataset constructed by Marx et al. (2020) that contains the results of the 2013 election at the polling center level.[15]

Electoral protests and exposure to conflict   Variables pertaining to violence and location of riots and protests come from the *Armed Conflict Location and Event Database* (ACLED), which provides details on the date, nature, and geo-location of conflicts (Raleigh

---

[14]Besides the small attrition in the sample, the aggregated results remain very close to official results given by the IEBC.

[15]The identification of the polling centers is similar for two elections, which enables the comparison between 2013 and 2017 outcomes at the polling center level.

et al., 2010). Based on the coordinates and the dates of violent episodes that occured in Kenya between 2007 and 2018, I retrieve the share of locations – in the ward or polling center surroundings – where such events took place, for different periods, and the number of associated fatalities. Appendix Table C.1 displays summary statistics at the ward level: on average, 18% of the wards experienced one reported conflict in the year preceding the election, usually in the form of a riot (81%). I also compute the distance from polling stations to the nearest episode of conflict that took place between the two rounds as a proxy for *direct* exposure to conflict.

2.2 Media and TV

Media access Today, the press in Kenya is considered *Partly Free* according to Freedom House (Freedom House, 2016). Freedom of the press is acknowledged by the Constitution and numerous media outlets are active and used to reporting a fairly diverse range of views. However, as noted by Freedom House, journalists routinely face threats and external pressure.[16] One of the main characteristics of media consumption is the heterogeneity in access to information. Radio is the most popular medium in Kenya, a pattern shared with the majority of other sub-Saharan countries (Nyabuga and Booker, 2013). Many radio stations broadcast locally in vernacular language and a large share of households own a radio set (Appendix Table C.2).[17]

In 2013, 43% of the territory was covered by the television signal. The country had relied on analog technology up to that point, and TV was widely available in large cities. Between 2013 and 2017, coverage increased substantially. This leap in signal availability was driven by Kenyan's commitment to the process of migration from analog to digital television, as part of the international agenda set by the International Telecommunication Union. The shift from analog to digital television makes it possible to reduce market tightness, since digital signals require less spectrum space, and to increase the number of available channels, allowing for broader media contents (Berger et al., 2010). As a result of massive government and ITU investments in television transmitters, more than 83% of the 7,003 sublocations had access to the technology in 2018. Hence, television availability has significantly increased in recent years. According to the Afrobarometer survey, around 40% of the population owns a TV set today, while more than half of the population stays informed via TV broadcasts at

---

[16]See Nyabola (2018) or Harwood et al. (2018) for more on political and financial pressures on the Kenyan media.

[17]Newspapers' reach is smaller: in 2016, around 25% of Kenyans read a newspaper at least a few times a week. In addition to limited English literacy rates in rural areas, which partly explain the limited consumption of print outlets, newspaper' circulation remains centered around urban areas because of distribution costs, a pattern that is shared with many other countries in the continent (Cagé, 2015).

Table 2.1: 2017 Presidential election results

	Mean	sd	Min	Max	N
2013 Election					
Registered voters in 2013	613	1,006	1	29,244	19,456
Turnout	0.88	0.11	0.00	1.00	19,456
Kenyatta's vote share	0.47	0.41	0.00	1.06	19,452
Odinga's vote share	0.46	0.39	0.00	1.81	19,452
2017 Election: First election					
Registered voters in 2017	801	1,226	1	30,572	21,015
Turnout in August	0.79	0.09	0.00	1.00	21,015
Kenyatta's vote share	0.01	0.39	-0.50	0.50	21,009
Odinga's vote share	0.47	0.38	0.00	1.00	21,009
2017 Election: Repeat election					
Turnout in October	0.36	0.32	0.00	0.94	21,015
Turnout in October (if open)	0.41	0.31	0.00	0.94	18,582
Kenyatta's vote share	0.96	0.08	0.00	1.00	18,375
Distance nearest protest	30.27	19.42	0.01	100.80	21,015

Notes: An observation is a polling center. **First election:** all polling centers are included (turnout: 79%). **Repeat election:** open stations refer to polling centers that were not blocked on the day of the second election (turnout 41% vs 36% when all observations are considered.) Distance to the nearest protests between the two 2017 elections is in kilometers.

least once a month (Appendix Table C.2). This gap between TV ownership and viewership is suggestive of the collective dimension in exposure to TV, which is often watched in public places.[18]

Television signal data Since 2013, data on TV coverage status are collected by the *Communications Authority of Kenya* (CAK) on a yearly basis, at the sublocation level. The organization is in charge of the implementation and follow-up of the digital migration plan: its missions include overseeing the construction of transmitters, granting licenses to television broadcasters, and measuring the current access to TV in the country.[19] To study how access to television affects political participation, I merge the annual records to compute TV penetration rates for each point of time between 2013 and 2018. While 3,192 sublocations

---

[18]Section 4.3 discusses the collective dimension in TV watching and emphasizes an increase in expenses in cafes and restaurants after the introduction of television signal.

[19]The digital migration plan in Kenya mostly consisted in the installation of terrestrial transmitters and the adoption of *Digital Video Broadcasting - Terrestrial* transmission standards (DVB-T and DVB-T2) (GSMA, 2016). Cable TV penetration is very limited in Kenya – and more generally in the East African region – mainly because of the installation and operational costs.

had access to TV signal in 2013; 5,142 were covered before the 2017 election. In other words, access to TV signal increased by 28% between the two presidential elections (Appendix Table C.3): the maps displayed in Figure 2.2 show that the signal spread significantly between 2013 and 2017. Notably, the capital Nairobi and its surroundings – as well as other large cities in Kenya such as Mombasa or Kisumu – already had access to TV in 2013. As new transmitters were installed across the country, signal availability spread and reached more rural and isolated areas. I also use the geolocalization of the TV transmitters active in 2017 provided by the CAK to compute the distance between each polling station and the closest transmitter. Table 2.2 shows summary statistics: the average distance to a transmitter was 53 kilometers in 2017. In Appendix Figure B.5, I illustrate the large variation in distance to the closest transmitter. In Section 3.1, I show that this variation in the distance and terrain barriers – measured as the polynomial of the distance and elevation statistics – are the main predictors of TV signal and make it possible to overcome endogeneity bias in TV reception. Between June 2017 and June 2018, the signal was extended to 690 new sublocations (Appendix Figure B.6). I use this set of polling centers – those were not exposed to TV broadcasts during the election, but benefited from the technology after the election –, as a robustness check in section 3.4.

TV broadcasts   To document the content of television news programs, I use data scraped from the website *KenyaMoja*, an online platform that collects the headlines of news articles or videos published or broadcast by the main national outlets. In the case of TV programs, the website gathers the titles of news stories broadcast on the news bulletins of the six most popular stations since 2012. Appendix Figure B.7 shows an example of a news story as available on the website. I construct a dataset that contains the headlines for the year 2017: for each headline (or news item), the dataset include the story's full title, the channel and the date of airing. It contains 86,189 new stories. Appendix Table C.4 shows summary statistics and the breakdown of news stories by channel is presented in Appendix Figure B.8.
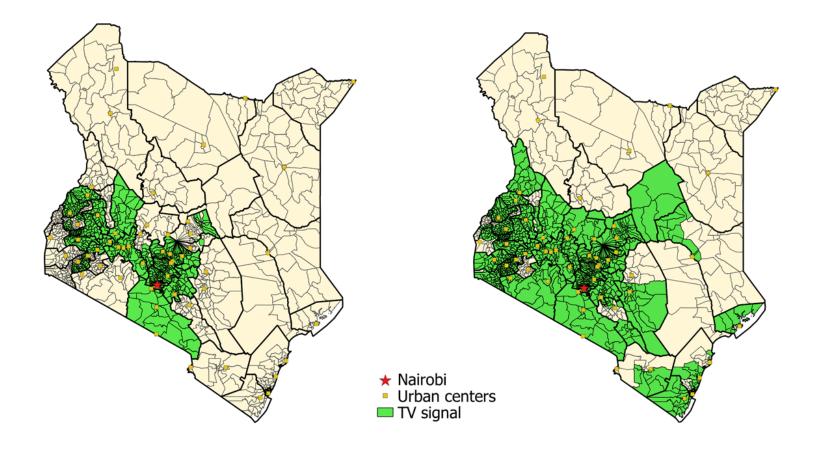
2.3   Administrative data and terrain characteristics

Administrative and socio economic data   Based on the geolocalization of the polling stations, as illustrated in Appendix Figure B.9, I match the electoral centers with the administrative entity to which they belong and their respective landmarks (county, constituency and ward level). I include a set of indicators in the dataset to account for community and ethnic belongings at a local level. I use data on ethnicity from the last *Demographic and Health Survey* (DHS) conducted in Kenya in 2014: based on the geocoded survey clusters, I retrieve the main ethnic group at the constituency level and construct an indicator of het-

Table 2.2: Summary statistics

	Mean	sd	Min	Max	N
TV status					
Analog in 2013	0.42	0.49	0.00	1.00	21,021
TV in 2017	0.71	0.45	0.00	1.00	21,021
Distance and terrain					
Distance transmitter (km)	52.65	68.75	0.14	548.09	21,021
Distance to capital (km)	214.90	117.92	0.20	730.47	21,021
Distance to the coast (km)	458.49	179.74	0.05	920.47	21,021
Mean altitude (m)	1,403.71	597.83	2.06	3,490.23	20,997
Std altitude (m)	35.42	43.66	0.47	652.02	20,997
Polling center area					
Area (km2)	2.55	10.06	0.00	364.36	21,021
Population (thousands)	27.36	9.46	1.05	140.32	21,021
Ethnicity					
# Ethnic groups > .1 of pop	2.99	1.41	1.00	5.00	21,011
Share incumbent and allies coethnic	0.31	0.46	0.00	1.00	21,021
Share opposition and allies coethnic	0.39	0.49	0.00	1.00	21,021
Socio-economic covariates					
Mean Light at night	1.80	6.38	0.00	63.00	19,514
Work for pay	18.41	10.79	0.30	74.60	21,021
Informal employement	12.14	5.95	1.30	98.60	21,021
Primary education	54.43	13.44	5.13	68.12	21,021
Secondary education	19.12	9.95	0.78	63.00	21,021

Notes: An observation is a polling station. Population comes from the 2019 national census. Geographic and elevation variables (area, distances, altitude, light at night) are computed on QGIS (author's calculations). Ethnicity variables come from the geolocalized *Demographic and Health Survey* of 2014. Socio-economic variables (education and employment statistics) come from the 2016 *Society for International Development* County Reports.

Notes: This figure shows the expansion of the TV signal in Kenya between 2013 (left-hand side map) and 2017 (right-hand side map) at the ward level.

Figure 2.2: Access to TV between 2013 and 2017

erogeneity based on the shares of the three main communities in the area (Table 2.2). I also use the geocoded DHS surveys from 2008 and 2014 to add economic and social covariates such as employment, consumption, or education statistics. Finally, I include satellite data on luminosity at night as a proxy for economic activity (Henderson et al., 2009).

Terrain features I add a set of spatial indicators at the polling center level. For each observation, I construct a Voronoi polygon: this approach allows me to connect each point on the map to the closest polling center so that the descriptive statistics derived at the center level can be considered as the average characteristics of all the places falling within its area of influence. The surface of the Voronoi polygons provides a fair estimate of population density. To approximate the inclusion of polling centers into urban networks and the strength of their linkages with relevant economic and political centers, I compute the distances to Nairobi, to the coast, and to the closest town with over 30,000 inhabitants. This exercise reveals the wide heterogeneity in urbanization schemes (Table 2.2). Moreover, I construct elevation indicators at the polling center and ward levels such as mean altitude and standard deviation of altitude in the area considered.

The final dataset includes fine-grained geographic data at the polling station level and contains detailed information on media access and socio-economic covariates. In the next section, I use the richness of the dataset to study the determinants of TV signal reception.

3   Empirical strategy and main results

In this section, I present the identification strategy and main findings. The main challenge when estimating the impact of TV on political outcomes stems from endogeneity in signal reception. To measure the real influence of exposure to broadcasts, we need to rely on the assumption that locations with or without access to TV in 2017 are similar in unobserved characteristics, conditional on observables. Importantly, controlling for observables, signal should not differ significantly across areas with different political preferences before its introduction. My empirical strategy proceeds in three steps. First, I investigate the main determinants of TV reception using topographic data and socio economic controls to address endogeneity concerns (subsection 3.1). Second, I conduct a series of exogeneity checks to test for the validity of the identification assumption, i.e. that signal availability is not significantly correlated with covariates that also influence political participation, conditional on observables (subsection 3.2). Third, I estimate a difference-in-differences model to estimate the impact of access to TV on turnout in both the first and the recall election: I study separately the effect of gaining access to TV between 2013 and 2017 on the difference in

turnout between the last presidential election and the two 2017 ballots (subsection 3.3). The results are corroborated by a series of robustness checks (subsection 3.4).

3.1 Determinants of signal availability

According to the signal propagation law of physics, the average reach of a television transmitter is between 60 and 100km. The likelihood of receiving a signal depends on the strength of the transmitter, the distance between the transmitter and the receiver and whether the receiver is in line of sight to the transmitter. In the literature, the location of transmitters and the signal strength have often been used to predict reception, or the theoretical signal availability in the absence of any terrain barriers, at a very granular level (see among others Olken, 2009; Enikolopov et al., 2011; Yanagizawa-Drott, 2014; Durante et al., 2019). I implement a similar approach and use the location of transmitters and terrain ruggedness data to predict signal availability: my identification strategy relies on variation in the distance to the transmitter and terrain ruggedness to exhibit exogenous variation in signal reception. As an illustration of the approach, Appendix Figure B.5 draws a straight line between each polling station and the closest transmitter. Out of the polling centers located in the 60-100 km radius around a transmitter, i.e. the standard interval of reach, 48% actually received the signal in 2017.

I now analyze more formally the factors that predict the likelihood of receiving the signal. The data provided by the CAK documents the status of signal reception at the sublocation level in 2013 and 2017, which allows me to have a thorough mapping of signal expansion over the period. To exhibit the main variables that explain signal propagation over time, I estimate the following model:

$$TV_{cwr} = \alpha + \delta.distance_{cwr} + \eta.topography_{cwr} + \kappa.urban_{cwr} + \mu.eco_{cwr} + \phi + \epsilon_{cwr} \qquad (3.1)$$

where TV_{cwr} is a binary variable equal to 1 if the polling station was granted access to TV between 2013 and 2017 in polling center c, ward w and county r; $distance_{cwr}$ stands for the distance between the polling center and the closest transmitter as well and the fourth polynomial of the distance, $topography_{cwr}$ contains measures of terrain ruggedness (mean and standard deviation of altitude at the ward and polling center level), $urban_{cwr}$ includes demographic and urbanization statistics, the surface of the polling center area and of the ward, as well as a set of distance variables – from the polling station to Nairobi, to the coast, or to the closest city (over 30,000 inhabitants), or to minor urban areas (above 1,000 people) –, and eco_{cwr} is a set of socio-economic controls at the ward level. I also include transmitter fixed effects ϕ. Standard errors are clustered at the closest transmitter level.

The results of model 3.1 are shown in Table 3.1. Column 1 suggests that within the area of a transmitter reach, a 1-km increase in the distance to the transmitter decreases the likelihood of receiving TV by 63%. The coefficient is significant at the 1% level. The inclusion of topographic controls in column 2 has little effect on the magnitude of the estimates (-60%), or on the adjusted R^2, suggesting that the polynomial of the distance to the transmitter controls explain most of the variation in TV reception. In columns 3 and 4, I include location controls and the estimates remain close to the first specification. This pattern is in line with existing results in the literature (e.g. Olken, 2009; Enikolopov et al., 2011) which use the polynomial of the distance to the transmitter as the main determinants of signal reception. These results are robust to a series of alternative specifications shown in Appendix Table C.5. I first include county fixed in place of transmitter fixed effects (columns 2, 3, 6, 8 and 10): the effect of distance to the transmitter holds within-country and is smaller in magnitude (-37%). The results are also robust to clustering standard errors are at the ward level, which is the smallest administrative unit within which I observe variation in TV reception (columns 3 to 6). Further, I exclude from the sample the polling centers that had access to analog TV before digital migration was implemented and the installation of new transmitters were installed from 2013.[20] The estimates shown in columns 7 and 10 of Appendix Table C.5 are very similar in magnitude to those displayed in Table 3.1 and significant at the 1% level. This indicates that the effect of distance to the transmitter, conditional on observables, is stable across polling stations with different access to the media before the investment plan in digital infrastructure.

3.2   Exogeneity checks

Table 3.1 brings to light observable determinants that explain whether or not a location got access to TV between 2013 and 2017. However, the fact that unobservable factors also significantly correlate with both political motives and signal reception may be a threat to the identification strategy. The following paragraph provides evidence on the credibility of the identification assumption: I run a series of placebo tests to ascertain exogeneity in signal reception conditional on observables. The first checks to be performed are on turnout and *ex ante* vote shares, since a significant correlation between access to the signal after 2013 and political outcomes before its introduction would violate the identification assumption. *Ex ante* differences in socio-economic characteristics also need to be examined to ensure that conditional on the distance to the transmitter and the set of controls, TV signal availability

---

[20]The roll out plan of digital migration focused on granting access to digital technology to regions previously covered by the analog signal (around one third of the territory) which may introduce a bias in the location choice of the transmitters.

Table 3.1: Predictors of TV signal availability

| | TV in 2017 | | | |
	(1)	(2)	(3)	(4)
Distance transmitter (km)	-0.631***	-0.598***	-0.469**	-0.501**
	(0.156)	(0.153)	(0.224)	(0.240)
Mean altitude (m)		-0.279**	-0.134	-0.138
		(0.113)	(0.125)	(0.122)
Std altitude (m)		-0.005	-0.002	-0.003
		(0.010)	(0.006)	(0.006)
Mean altitude ward		0.071	0.096	0.090
		(0.082)	(0.069)	(0.067)
Std altitude ward		-0.017	-0.021	-0.019
		(0.026)	(0.023)	(0.022)
Distance to capital (km)			0.159	0.034
			(0.514)	(0.569)
Distance to the coast (km)			0.413	0.299
			(0.580)	(0.554)
Dist to closest town (km)			-0.186	-0.185
			(0.305)	(0.307)
Observations	19,502	19,502	19,502	19,502
Mean DepVar	0.70	0.70	0.70	0.70
Sd DepVar	0.46	0.46	0.46	0.46
Adjusted R2	0.36	0.37	0.40	0.40
Transmitter FE	✓	✓	✓	✓
Transmitter polynomial	✓	✓	✓	✓
Topography		✓	✓	✓
Location controls			✓	✓
Pop-Eco controls				✓

Notes: Standard errors in parentheses are clustered at the closest transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is a polling center. Transmitter polynomial include the fourth polynomial of the distance between the polling center and the closest transmitter. Topography include altitude and the standard deviation of altitude in the Voronoi polygon of the center, as well as similar statistics computed at the ward level. Location controls include distance to the capital, to the coast, to the closest town above 30,000 inhabitants, to the closest urban center, as well as the squared distances. Pop-Eco controls include population in the ward, density, light at night in the ward (mean and standard deviation).

is not correlated with previous differences in covariates.

I estimate the following model:

$$z_{cwr} = \alpha + \beta.TV_{cwr} + \delta.distance_{cwr} + \eta.topography_{cwr} + \kappa.urban_{cwr}$$
$$+\mu.eco_{cwr} + \phi + \gamma + \epsilon_{cwr} \tag{3.2}$$

where z_{cwr} is a dependent variable considered for a polling center c, in ward w and county r, such as past electoral results, or economic and demographic characteristics. TV_{cwr} stands for TV coverage status in 2017. I control for the propagation controls estimated in model 3.1: namely the fourth polynomial of the distance to the closest transmitter $distance_{cwr}$, topographic controls $topography_{cwr}$, urbanization status $urban_{cwr}$, and socio-economic covariates eco_{cwr}. I include closest transmitter fixed effects ϕ and county fixed effects γ. Standard errors are clustered at the closest transmitter level.

Table 3.2 shows the results of equation 3.2 where the dependent variables are 2013 political outcomes or community belongings. The estimates suggest that there is no significant correlation between the covariates z_{cwr} and gaining access to TV between 2013 and 2017 conditional on signal predictors identified in paragraph 3.1 : the $\beta = 0$ result supports the assumption that the introduction of the signal after 2013 can be considered exogenous to previous political outcomes (columns 1 to 3). It is worth noting that without including these propagation controls, the correlation between turnout in 2013 and TV is negative and significant at the 5% level, as shown in Appendix Table C.7. In columns 4, 5 and 6, I estimate model 3.2 with an indicator of ethnic homogeneity or the share of incumbent (or of the opposition) co-ethnics in the ward as dependent variables. This is an important test to carry out because of the substantial links between political leaders and their communities (Burgess et al., 2015; Horowitz and Long, 2016). Here also, the estimates are close to zero and not significant, which suggests an absence of ethnicity bias in signal reception.[21] Another possible source of endogeneity is that differences in socio-economic characteristics could be correlated with both political behavior in 2017 and signal reception. I run a similar exogeneity test using as outcomes labor, education or consumption statistics from survey data and show the results in Appendix Table C.8.[22] Conditional on the propagation controls, I do not find a significant correlation between TV and work status (columns 1 and 2), nor with the education status of individuals (columns 3 and 4), and all the coefficients are close to

---

[21]The results of regression 3.2 without controlling for signal predictors displayed in Appendix C.7 suggest that the places with a higher share of incumbent pro-ethnic were significantly less likely to get accessed to the signal after 2013.

[22]Employment and education status are from the 2014 Demographic and Health Survey and I use the 2016 Kenyan Integrated Household Budget Survey (KIHBS) for household and consumption data.

zero.[23] In Table Appendix C.9, I look at *ex ante* differences using data from the 2008 round of the DHS and find no evidence of pre-trends in the allocation of the signal conditional on observables. I also conduct a similar robustness check as in paragraph 3.1 and estimate model 3.2 on the set of polling stations without access to analog TV in 2013. The results are shown in Appendix Table C.10 and C.11 and are very similar to the estimates found using the whole sample: the coefficients are close to zero and not significant.

Hence, these findings imply that distance to the transmitter polynomial and geographic controls are satisfactory predictors of signal availability and that more importantly, they make it possible to overcome endogeneity threats due to non-randomness in transmitter location. In other words, it gives credit to the assumption that we can estimate the causal effect of exposure to TV on political outcomes in 2017.

3.3   TV and turnout in the repeat election

In order to test whether reception of a TV signal had an effect on turnout in either the first election of August or the recall election of October, I estimate a difference-in-differences model, in which I consider the difference between turnout in 2013 and 2017, and define the treatment group as the polling stations that gained access to TV between the two presidential elections. I estimate the following regression:

$$
\begin{aligned}
turnout2017_{cwr} = \alpha + \mu.turnout2013_{cwr} + \beta_1.TV_{cwr} + \beta_2.Incumbent2013_{cwr}+ \\
\beta_3.TV_{cwr} \times Incumbent2013_{cwr} + \rho.signal_{cwr} + \phi + \gamma + u_{cwr}
\end{aligned}
\tag{3.3}
$$

where $turnout2017_{cwr}$ stands respectively for turnout in 2017, in the first or the recall election in polling center c, ward w, and county r, and $turnout2013_{cwr}$ is turnout in 2013.[24] TV_{cwr} is an indicator variable that takes the value 1 if the polling center began receiving the signal between 2013 and 2017. $Incumbent2013_{cwr}$ is either a categorical variable equal to 1 in polling stations where the incumbent President Kenyatta's vote share in 2013 was above 50%; or a continuous variable standing for the vote share of the incumbent in the previous election, centered around 0. $signal_{cwr}$ stands for the propagation controls such as distance to the transmitter polynomials, topography and urbanization covariates, described in model 3.1. I include closest transmitter fixed effects ϕ and county fixed effects γ. Standard errors are clustered at the closest transmitter level. First, I estimate a baseline model with

---

[23]In Appendix Table C.6, I show the raw correlations between TV availability and other covariates: access to TV is positively correlated with living standards or education level.

[24]This specification is equivalent to using the difference between 2017 and 2013 $\Delta turnout$ as the dependent variable. I include turnout in 2013 on the right-hand side of the equation for a reasons of readability of the estimates.

Table 3.2: Exogeneity checks: political and community covariates

	2013 election			Community			
	(1) Turnout in 2013	(2) Incumbent score	(3) Opposition score	(4) Share 1st ethnic group	(5) Incumbent coethnic	(6) Opposition coethnic	(7) Violence 2007-2013
TV in 2017	-0.005	0.009	-0.016	0.018	-0.009	-0.010	-0.029
	(0.004)	(0.012)	(0.014)	(0.195)	(0.012)	(0.027)	(0.032)
Observations	18,022	18,022	18,022	18,022	18,022	18,022	18,022
Mean DepVar	0.88	0.47	0.46	3.00	0.32	0.37	0.26
Sd DepVar	0.11	0.41	0.39	1.41	0.47	0.48	0.44
Adjusted R2	0.27	0.83	0.82	0.57	0.89	0.90	0.19
County FE	✓	✓	✓	✓	✓	✓	✓
Transmitter FE	✓	✓	✓	✓	✓	✓	✓
Propagation controls	✓	✓	✓	✓	✓	✓	✓

Notes Standard errors in parentheses are clustered at the transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is a polling station. TV in 2017 is a binary variable equal to 1 if TV is available in the sublocation. Share first ethnic group is the quantile of the share of the dominant ethnic group in the ward as a proxy for the degree of ethnic homogeneity. Incumbent (resp Opposition) co-ethnic is a binary variable equal to 1 if the dominant group in the ward is part of Kenyatta's (resp Odinga) political coalitions. Violence 2007-2013 is equal to one when at least one episode of violence is recorded in the ward by ACLED across the period.

the coefficients $Incumbent2013$ and TV_{cwr} as two separate explanatory variables.[25] I then interact them in the following specifications to investigate the possible heterogeneous effect of TV across political camps.

Results of model 3.3 are shown in Table 3.3. In columns 1 to 3, the dependent variable is turnout in the August 2017 election, and I consider the recall election of October as outcome in columns 4 to 6. First, I find that access to TV in 2017 did not affect turnout in the first – *regular* – election: the estimates from the baseline and interaction specifications are close to zero and not significant. Moreover, places that voted more heavily in favor of the incumbent in 2013 were significantly more likely to cast a ballot in 2017: a 1-percent increase in the incumbent's vote share in 2013 increases turnout in the polling station by 0.03 percentage points (column 1). Yet, I do not find heterogeneous effects of exposure to TV across political parties in columns 2 and 3. [26]

Turning to the recall election of October, the estimates displayed in columns 4 to 6 suggest a very different pattern. First, I find that the effect of TV on turnout is negative (column 4): as well as being insignificant at any conventional level, the size of the coefficient (-0.09) differs from the zero result of column 1. More interestingly, columns 5 and 6 suggest that exposure to TV has a heterogeneous effect. While turnout dropped in pro-opposition

---

[25]In this case, the estimated equation is: $turnout2017_{cwr} = \alpha + \mu.turnout2013_{cwr} + \beta_1.TV_{cwr} + \beta_2.Incumbent2013_{cwr} + \rho.signal_{cwr} + \phi + \gamma + u_{cwr}$.

[26]In Appendix Table C.12, I find that the results on the first election are very similar when considering the polling stations which were open in both rounds of the 2017 election.

Table 3.3: Access to TV and turnout

	Turnout in August			Turnout in October		
	(1)	(2)	(3)	(4)	(5)	(6)
TV in 2017	0.000	-0.002	0.003	-0.009	-0.038**	0.022
	(0.005)	(0.004)	(0.006)	(0.007)	(0.015)	(0.018)
Incumbent vote Sh. in 2013	0.027**		0.020	0.443***		0.355***
	(0.013)		(0.012)	(0.061)		(0.064)
Pro Incumbent 2013		0.008			0.168***	
		(0.005)			(0.035)	
TV 2017 × Pro Incumbent 2013		0.009			0.102***	
		(0.007)			(0.026)	
TV 2017 × Incumbent Vote Sh. 2013			0.012			0.150***
			(0.012)			(0.043)
Observations	18,000	18,000	18,000	15,946	15,946	15,946
Mean DepVar	0.79	0.79	0.79	0.41	0.41	0.41
Sd DepVar	0.09	0.09	0.09	0.31	0.31	0.31
Adjusted R2	0.52	0.51	0.52	0.91	0.89	0.92
County FE	✓	✓	✓	✓	✓	✓

Notes: Standard errors in parentheses are clustered at the closest transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is polling station where the two rounds took place (first and recall election). Pro Incumbent 2013 is a binary variable equal to 1 if the incumbent Kenyatta ranked first in the 2013 election. Incumbent Vote Sh is continuous and centered around 0. In Columns 1 to 3 (resp. 4 to 6), the dependent variable in turnout in the first election of August (resp. turnout in the recall election of October).

polling stations by 4 percentage points (significant at the 5% level), it is 7 percent higher in polling centers that voted in favor of the incumbent in 2013 (significant at the 1% level). In columns 3 and 6, support for the incumbent is measured as a continuous variable, centered around 0.[27] The interaction term suggests that the effect of having access to TV grows stronger as the support for the incumbent in the previous election increases. A 1-percent increase in the score of the incumbent adds around 0.15 percentage point to the baseline effect of TV. All these results are robust to clustering the standard errors at the ward level, as shown in Appendix Table C.13.

To provide more evidence on the heterogeneous effect of TV, I estimate a model similar to equation 3.3, in which I replace the score of the incumbent in 2013 with his score in the August 2017 election, and use turnout in the recall election as the dependent variable. This approach is in the spirit of a first-difference model: it relies on the absence of significant correlation between TV and turnout in the first election documented in Table 3.3, which makes it possible to alleviate the risk of bias that would stem from a correlation between TV and turnout in the first election. The results are shown in Appendix Table C.14 and

---

[27]The variable *Incumbent*2013 vote share is equal to 0 when the two candidates get the exact same score in the first election.

are in line with the previous estimates: the average effect of TV on turnout is negative besides not significant (column 1), as was the case in the main specification. In line with the previous findings, exposure to TV significantly decreases turnout in polling stations where the opposition came out on top in the first election (-3pp) but we observe an increase where the incumbent ranked first in the August election (+3pp).

Figure 3.1 provides a graphical illustration of the results. It displays the predicted turnout in the repeat election in places with access to TV or without, when equation 3.3 is estimated with an interaction between access to TV and the incumbent's score in 2013, categorized in quantiles. In pro-opposition bastions, i.e. on the left side of the graph, TV has a negative effect on turnout but this effect is offset as the score of the incumbent increases. In the last quantiles of the distribution, access to TV amplified mobilization in the recall election.[28]
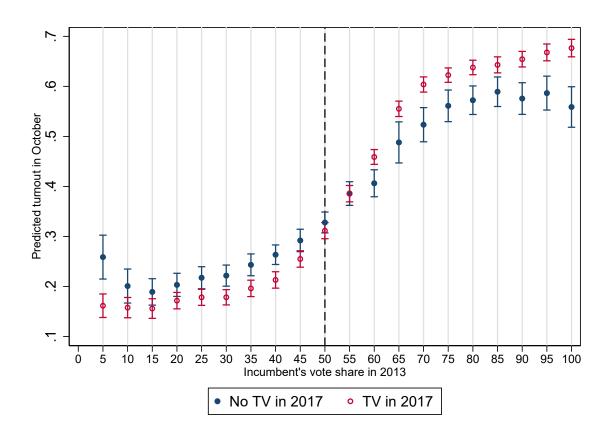
This set of results brings to light two main takeaways. First, the effect of exposure to TV was different in the August and the October polls. While exposure did not affect turnout in the *regular* election, it seems to have played a role in the recall one, suggesting that the electoral context may alter by the media's impact on voters. Second, the effect of being exposed to TV in a highly uncertain period is different across the political camps. Exposure to TV seems to have deterred voters from casting a new ballot where the opposition ranked first, whereas it amplified support for the incumbent where the latter was winning the race. In Section 4, I consider different possible mechanisms to rationalize these results, and explore how they can relate to the coverage of the crisis in the broadcasts.

3.4   Robustness checks

Results shown in Table 3.3 are robust to a series of alternative specifications. Appendix Table C.13 shows the results with standard errors clustered at the ward level. Appendix C.14 presents the results using a first-difference setting, i.e. defining political heterogeneity using the results from the first 2017 election. Moreover, in Appendix Table C.15, I interact signal determinants with an indicator variable equal to 1 in pro-incumbent polling stations to further address possible bias between the treatment determinants and political preferences and find very similar results.

I now present two additional tests that provide further empirical support for the previous findings. First, I conduct a placebo test and show that TV has no effect in places that were granted access to the signal after the 2017 election, which confirms that the results are not spurious. Second, I restrict the sample to polling stations that had no access to TV in 2016

---

[28]Appendix Figure B.10 shows a similar graph for the August election and further illustrates the null impact of exposure to TV in the first election.

Notes: The figure displays the predicted turnout in October in polling stations covered by the TV signal vs. those not covered. Turnout is predicted for each quantile of Kenyatta's score in the first election: for instance, the coefficients at x=100 are estimated on polling stations where the incumbent score in August was above the 95th threshold. The reference category is the first quantile (5% of polling stations where Kenyatta obtained his lowest score in August).

Figure 3.1: Marginal impact of TV on turnout in the repeated election of October

in order to exclude the more urban areas and test for the validity of the results in places that have been exposed to TV for a shorter period of time.

TV signal in 2018   First, I look at polling stations that were gained access to a TV signal between August 2017 and August 2018 to check that the previous findings are not spurious. Despite the fact that some centers in this subgroup may have begun receiving a signal before October 2017, implying that they were covered by the signal during both rounds of the election[29], the majority of the observations benefited from the introduction of digital broadcasting only in the months after the election. Appendix Figure B.6 shows the areas that gained access to a digital signal during this period. In total, 690 sublocations were given access to a signal during this period, which corresponds to approximately 3,000 voting centers. Given the staggered introduction of digital television in the country, it is plausible that the areas that gained access in 2018 could have been part of the 2017 or 2016 beneficiaries if no delays had occured after 2013.

I estimate model 3.3 on the sample of polling stations without a signal in 2017 and regress turnout on access to TV in 2018. The results are shown in Appendix Table C.16. Neither the coefficient *TV in 2018* nor the interaction term is significant. This result supports the previous finding: when deciding whether or not to mobilize, voters were influenced by their exposure to TV in areas that received the signal before the election, while the later introduction of digital technology had no impact on remobilization. Moreover, the role played by the incumbent margin in the first round is still significant at the 1% level and the magnitude of the estimates for both the August and the October elections is consistent with the results found in Table 3.3. This pattern reflects a homogeneous elasticity of remobilization with respect to the incumbent's past vote share.

Polling centers without TV in 2016   I also replicate the analysis on the subset of polling centers that had no access to a TV signal in 2016 (5,663 polling stations). 32% of them benefited from the expansion of the digital signal before the 2017 election. The results are displayed in Appendix Table C.17. Arguably, households in regions that gained access to TV between 2013 and 2017 may also have increased their Internet consumption over the period, and there could be concern over whether or not the treatment effect might be contaminated by the simultaneous expansion of the Internet in the country.[30] Given the

---

[29]This caveat is due to the structure of the CAK data, which is collected on a yearly basis starting in August.

[30]Previous studies have shown the effects of the increase in Internet penetration on outcomes such as perceptions of corruption (Guriev et al., 2019) or employment (Hjort and Poulsen, 2019), which could possibly correlate with voting outcomes.

high prices of data bundles and fixed broadband connections in Kenya, but more generally in sub-Saharan African countries, this would be a possible concern with regard to more urban, wealthy and well connected parts of the country. Another concern stems from the fact that people's exposure to the crisisin these more populated areas was based on direct experience of the protests or election events rather than on TV broadcasts. Hence, focusing on peripheral areas that are located far from the crisis hotspots and therefore less likely to have high Internet penetration allows me to circumvent these two alternative channels. Indeed, in counties where less than 25% of the polling stations receive a television signal in 2016 [31], more than 90% of the inhabitants had not used the Internet in the three months before being interviewed.[32] Results are shown in Appendix Table C.17. In polling centers with no signal in 2016, access to TV broadcasts did not impact turnout in the August election but, in line with the results shown in Table 3.3, the effect of TV on pro-opposition and pro-incumbent voters was opposite (-5pp vs +10pp).

4   Mechanisms and Discussion

How can we explain the fact exposure to television did not affect turnout choices in the first election but did so heterogeneously in the recall election? This section provides further results and proposes some mechanisms to explain the findings. First, to understand voting behavior in the recall election, I examine TV coverage of the crisis, implementing text analysis methods (subsection 4.1). Second, I investigate the link between exposure to the crisis through TV and direct experience of the conflict (subsection 4.2). Last of all, I provide evidence of the collective dimension of TV exposure and argue that the results may be driven by social coordination mechanisms (subsection 4.3).

4.1   Content of TV broadcasts

To better understand the reason why exposure to TV had an impact on voters only in the repeat election, and the mechanisms explaining why it affected voters from the two camps differently, one should first remember that the second election took place in a climate of high uncertainty and during a period of exacerbated violence in the country (Appendix Figure B.3). With that in mind, this subsection examines the content of the broadcasts in order to

---

[31]The counties where less than 25% of the polling stations receive the TV signal in 2016 were: Bomet, Migori, Wajir, Baringo, Tana River, Garissa, Samburu, Trans Nzoia.

[32]This statistics comes from the 2016 *Kenyan Integrated Household Budget Survey*. These regions are also further from urban centers as the mean distance to a city with over 1 million inhabitants is 317 km, for a national average of 197 km.

document TV coverage of the electoral troubles. First, I look at the vocabulary and the most prevalent words used during the period under study. Using text analysis methods (Gentzkow et al., 2019), I provide evidence of the emphasis placed on violence and uncertainty in the broadcasts between the two elections. Second, I investigate the concentration of the news coverage in terms of diversity in reported topics.
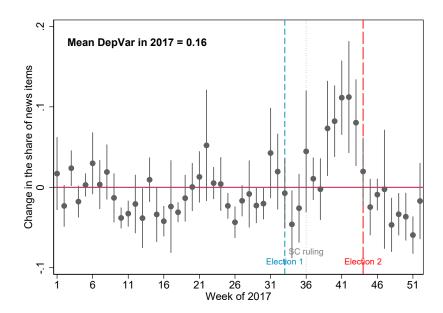
To do so, I create a dataset that gathers the headlines of news bulletins aired by the five most popular TV channels in Kenya in 2017 (Appendix Figure B.8). Appendix Table C.4 provides summary statistics on the headline dataset: it includes 86,189 headlines and 236 news items reported per day on average. The methodology to clean and format the data is described in Appendix A.1.[33] I construct then an indicator of conflict-related items. This indicator is based on a list of keywords belonging to a *crisis lexicon* – the list of tokens used is shown in Appendix Table A.1a, as well as an alternative and more restricted word set used to perform robustness checks in Appendix Table A.1b –, and allows me to measure the share of news stories related to political uncertainty or episodes of violence. On average, 16% of the items are conflict-related per day.

The goal is to examine whether this indicator varied during the Kenyan electoral crisis, and to estimate the magnitude of the emphasis placed on the troubles in the broadcasts. I aggregate the data at the week level and regress the share of conflict-related stories on week indicator variables, using as baseline category the mean of the share in 2017. The estimates are plotted in Figure 4.1. The coefficients suggest an increase in the coverage of conflict-related stories in the two weeks preceding the first election and in the weeks following its nullification by the Supreme Court. The magnitude of the increase is sizable; in the month preceding the recall election, TV broadcasts reported on electoral conflict around 9% more than on the average week. Notably, we do not observe such a pattern for the first election: while the share of conflict-related stories was higher than the average in the two weeks before the August election, the magnitude of the coefficients is smaller and, if we extend the window to the two months before the elections, we actuallt see a decrease in the share. These results are robust to an alternative list of keywords, for instance using a restricted set of tokens that explicitly relate to violence (Appendix Figure B.11) and to the inclusion of the 2018 headlines (Appendix Figure B.12).

I also conduct a *term frequency-inverse document frequency* (tf-idf) exercise where I define as a document the set of news stories broadcast by a channel during a given month and construct a corpus of tokens used by the five channels in 2017.[34] The final size of the

---

[33]Cleaning includes removing stopwords and punctuation or applying a translation algorithm from Swahili to English. Indeed, the large majority of the headlines are in English but it is common to find Swahili words in a sentence written in English (see Appendix A.1).

[34]Given the short length of the headlines, I define documents at the month level to have a larger set of

Notes: The figure plots the estimates retrieved from the regression of the share of conflict related stories per day on week fixed effect. The share of conflict related stories is computed using the wide set of keywords (see Appendix A.1a). The reference category is the average share of conflict related stories over the year i.e. 0.13. The green vertical line stands for the week of the first election (August 8th 2017), the red line for the week of the Supreme Court decision to nullify the election (September 1st 2017) and the blue line for the recall election (October 26th 2017).

Figure 4.1: Share of conflict-related content

corpus is 26,980 tokens. For each word, I compute a *tf-idf* score (see Appendix A.3 for the technical procedure): by construction, the score is larger for tokens that are frequent in one document but less so in others. In our setting, it allows to retrieve the words that are the most prevalent during a given month, in relative terms vis-a-vis the other months, and to infer the most topical news items at different points in time.[35] In Table Appendix Table C.18, I provide the list of the most prevalent tokens of 2017 at the monthly level, after removing the names of local places and individuals.

As expected, the most predictive words at the time of the electoral crisis relate to the

---

words and therefore less noisy results. Moreover, given that the dates of the two elections and the Supreme Court decision all fell at the beginning or the end of the month (August, 8, September 1, and October, 26, respectively), I can group observation by month and look at the preceding group as a proxy for the pre or post-electoral periods. I clean the headlines and reduce the dimensionality of the word matrix as a preliminary step, and remove from the analysis the tokens that are used in more than 33% and in less than 1% of the headlines. The threshold can be changed to be more or less restrictive but the following results remain unaffected by the selected boundaries.

[35]One difficulty arising from such a token prevalence analysis is that many stories which hit the front page during a short period of time pertain to events that have occurred in specific places and involved particular individuals: with no context in which to situate the protagonists, or ant detailed knowledge of the geography of Kenya, the readability of tokens related to very specific stories is limited.
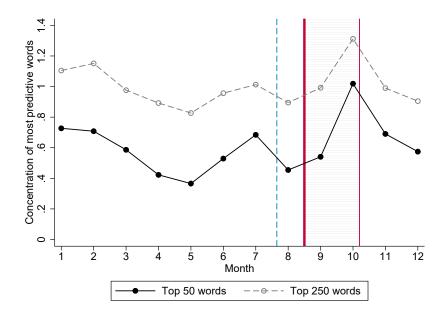
political situation. In September, the nullification of the election appears as the main story – the expression "the Maraga Decision" – which referred to the ruling pronounced by the President of the Supreme Court David Maraga – ranks first; the words "judgment", "dissent", "grandstand" or "storm" follow closely after. In October, i.e. the month of the boycott decision and of the demonstrations organized by the opposition, the most predictive word is "anti-IEBC", and tokens such as "commission", "toll", "participate", "riot" and "anti-riot" all appear among the 20 most predictive items.

To delve deeper into the news content and provide a more comprehensive analysis, I take a more agnostic approach by looking at the concentration of the most predictive words for each month of 2017. This strategy has two main advantages: first, it allows me to overcome the limitations due to context-specific references. More importantly, it also provides a broader picture of the coverage: the analysis of the *tf-idf* scores in a document period does not provide an accurate outlook of the relative amount of time spent airing stories related to those keywords. Therefore, I construct a dataset that includes the Z tokens with the highest scores – where Z is a fixed threshold – and compute a *concentration of coverage* index by summing the squares of the *tf-idf* scores of the top words used during a given month.[36] A larger index implies a greater predictive power of the set of the most predictive tokens, or in othe words, a more concentrated coverage, as the square of the words with the greatest scores (i.e. the most predictive words) has an increasing effect on the concentration index, in the spirit of a Herfindhal-Hirschman score. Conversely, if in a certain month, the Z most prevalent tokens have lower *tf-idf* scores, the concentration index will be smaller, implying a more general or more diverse news coverage over the period.

Figure 4.2 displays the concentration index for the year 2017. The dotted blue line stands for the first election date – August, 8, 2017 –, while the dates of the nullification and the recall elections – September, 1, 2017 and October, 26, 2017 – are drawn in red. Strikingly, we observe a stark peak in October, i.e. the month that preceded the recall election, meaning that the coverage was the most concentrated or the least diverse in the year 2017. This result holds when changing the number of prevalent tokens in the score computation (Z=50 or Z=250). Notably, concentration was also high in July, which was the month before the first election, and at the beginning of the year, when the registration of voters took place. This drop in news diversity before the first and, more importantly, the recall election is in line with a narrative of an increased focus on political issues in the country before the

---

[36]For each month $m \in \mathbb{M} = \{1, \ldots M\}$, I compute the following index $score_{it}$, i.e. the score of word i in month t:

$$concentration_m = \sum_{\substack{i \in \{1, \ldots, N\}, \\ t \in \mathbb{M}}} (score_{it})^2 . \mathbb{1}\{t = m\}$$

.

Figure 4.2: Concentration index of top prevalent tokens in 2017

elections, notably through a tightening of the news coverage on TV.[37]

This greater emphasis on the electoral crisis and the period of conflict on TV, as well as the drop in news diversity in the broadcasts can explain why TV had an unusual effect on voters during the recall election. With regard to heterogeneity, one plausible mechanism could be that exposure to the news resulted in an increased willingness to show support for the leader whose August victory had been nullified by the Supreme Court in pro-incumbent areas. Conversely, in pro-opposition areas, the incumbent's supporters may have been deterred from casting a ballot because of a higher perceived risk. In the next paragraph, I delve into this channel and relate exposure to TV on the one hand to direct exposure to the post-nullification protests on the other.

---

[37]In Appendix Figure B.13, I give further evidence of the difference in news coverage across channels: it plots the concentration index computed separately for the four most popular private channels in Kenya (Citizen TV, NTV, KTN, and K24) and the state channel KBC (Kenyan Broadcasting Corporation). While we observe a higher concentration index before the August election on the public channel, the graph suggests that the peak in concentration before the recall election shown in Figure 4.2 was driven by the private channels.

4.2 Direct and indirect exposure to violence

The emphasis placed on the institutional crisis on TV and the use of the lexicon of political conflict raise the question of the possible effect of being directly exposed to the electoral protests and riots. Indeed, if the coverage of the events on TV – which can be considered as *indirect* exposure to the crisis – had an impact on voters, then *direct* exposure may also have influenced voting choices. Moreover, a subsequent question arises with the potential interaction between *direct* and *indirect* exposure to TV. To investigate these possible channels, I use geocoded data on the location of protests and violent outbursts between the two rounds, taken from the *Armed Conflict Location & Event Data* (ACLED), as a measure of *direct* exposure to conflict. I use the number of politically related protests or conflicts that took place in polling station' wards between the two polls, and compute the distance between each polling station and the nearest protest episode. Appendix Table C.1 shows descriptive statistics on past episodes of political protests at the ward level.

I estimate an augmented version of model 3.3 in which I replace the dichotomous treatment of gaining access to TV over the period by three distinct treatment variables. I separate polling stations which have been only exposed to conflict through *direct* exposure – i.e. if a protest occurred in the wards between the two rounds –, from those whose exposure was only *indirect*, i.e. via TV broadcast. The last treatment group includes the polling centers characterized by both *direct* and *indirect* exposure to the conflict, such that the polling stations without access to TV and with no conflict form the control group. I estimate treatment effects using the following regression:

$$turnout2017_{cwr} = \alpha + \mu.turnout2013_{cwr} + \eta_1.Violence_{wr} + \eta_2.TV_{cwr} + \eta_3.Violence_{wr} \times TV_{cwr} +$$
$$\kappa_1.Violence_{wr} \times Inc_{cwr} + \kappa_2.TV_{cwr} \times Inc_{cwr} + \kappa_3.Violence_{wr} \times TV_{cwr} \times Inc_{cwr} +$$
$$+ \rho.signal_{cwr} + \phi + \gamma + u_{cwr} \quad (3.4)$$

where $turnout2017_{cwr}$ stands for turnout in the repeat election in polling center c, ward w, and county r, and $turnout2013_{cwr}$ is turnout in 2013. TV_{cwr} is an indicator variable that takes the value 1 if the polling center began receiving the signal between 2013 and 2017. $Violence_{wr}$ is equal to 1 if at least one protest or riot occurred in the ward between the rounds. Similarly to model 3.3, I interact the treatment variable with a partisanship indicator variable Inc_{cwr}, equal to 1 when the incumbent ranked first in the last election of 2013 in the polling center. $signal_{cwr}$ are propagation controls. I include closest transmitter fixed effects ϕ and county fixed effects γ. Standard errors are clustered at the closest transmitter level.

240

The detailed results are shown in Appendix Table C.19 and I plot the coefficients η and κ in Figure 4.3. The estimates from the baseline model are shown with blue circles: turnout was not affected in polling stations without access to TV but directly exposed to conflict. The indirect effect of exposure is small and not significant, but the combination of both direct and indirect exposure has a significant and negative effect on turnout (-2.3pp). Nevertheless, similarly to the previous findings, these coefficients hide a wide heterogeneity across political camps. While I do not find evidence of a different effect of direct exposure between pro-opposition (yellow diamonds) and pro-incumbent (red squares) polling stations, the impact of *indirect* and *direct $\times$ indirect* exposure are of opposite signs and larger in magnitude. In polling centers that voted in favor of the opposition in 2013, indirect exposure to conflict reduced turnout by 4% (significant at the 5% level), and the effect of direct and indirect exposure is of similar magnitude. Conversely, we find a positive effect of both treatments (*indirect* and *indirect $\times$ direct*) in polling centers where the incumbent won in 2013 (+10%).[38] Overall, these findings suggest that exposure to protests does have an effect on turnout only through the medium of TV.

However, this approach has a major caveat since the location of the protests may not be random across the territory. The estimates of model 3.4 are likely to be biased as it is plausible that voting choices are singular in places where riots took place.[39] To better address this endogeneity concern, I use the distance to the nearest protest as a measure of exposure to conflict and study how an increase in distance affects turnout. This strategy relies on the assumption that because of the location of the protests, which often occur in large avenues or city centers, a marginal increase in the distance between the crisis hotspots and the location of the polling center could be considered as exogenous. Therefore, I estimate an alternative version of model 3.4, in which $Violence_{wr}$ is replaced by the distance between the polling center and the nearest protest between the two rounds.[40]

Appendix Table C.20 shows the results. Considering all polling stations together, none of the treatment variables has an effect on turnout (columns 1 and 3 display estimates close to zero and not significant). In line with the previous findings, this null effect seems to stem from the opposite signs of the estimates on the two sides of the political spectrum. In polling stations that voted mostly for the opposition in 2013, exposure to TV decreases turnout (-2.5pp), while a 1-kilometer increase in the distance to the nearest protest has a positive effect

---

[38]The difference between the two coefficients is not statistically significant in both pro-opposition and pro-incumbent polling centers as shown in Appendix Table C.19 (p-values of 0.66 and 0.20).

[39]The sense of the bias would depend on whether voting and protest behavior are complement or substitutes: in the first case, the effect would be overestimated while we would underestimate the impact of conflict on turnout.

[40]We therefore move from a treatment defined at the ward level to one defined at the polling-station level.
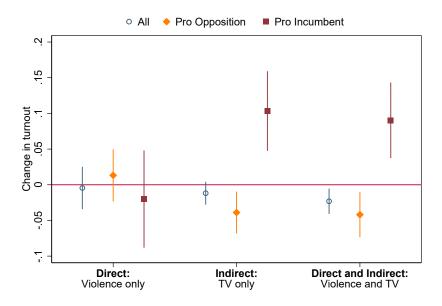
on turnout (+0.6pp). The intermediation of TV seems to dampen the effect of distance: the estimate $TV \times distance$ is negative and statistically significant. These results are consistent with the following narrative: direct exposure to conflict has a deterring effect in areas that voted mostly in favor of the opposition, and voters tend to participate more when they are further away from the riots. Exposure to TV counters the positive effect of distance by making the protests seem closer to voters. With regard to the incumbent's supporters, the effect of being exposed to the protests is flipped: turnout is higher in places covered by the TV signal (+7pp), and distance has a negative effect. It suggests that voters who found themselves closer to an episode of conflict were more likely to vote. The protests were often launched by opposition supporters calling for the resignation of the electoral commission, and so going to the polling stations in greater numbers may have been perceived by the incumbent's electorate as a way to express their support. However, the effect of distance is also offset by reception of a TV signal. Considered together, these findings suggest that the reaction to nearby protests and to the broadcasts are different between the two camps but that the latter counters the former in both of them. In Appendix Table C.21, I exclude the polling centers from the three main cities of the country (i.e. Nairobi, Mombasa and Kisumu), as the density of the riots was much higher, and find very similar results.

Since the previous findings suggest that the interaction between direct and indirect exposure to conflict matters, I explore whether long-term experience of violence also affects voters. To do so, I split the sample across wards where at least one political protest occurred between the last two presidential elections and wards where none did, i.e. between 2013 and 2017, and estimate model 3.3 separately on these two groups. The results are shown in Appendix Table C.22: polling stations exposed to political unrest in the past do not display different voting patterns from the others.[41]

4.3   Collective exposure to TV

Finally, voters' reactions to the broadcasts may have been strengthened by the collective dimension of exposure to TV. As underlined in Section 2.2, TV ownership in Kenya is lower than TV consumption, and TV is often watched in public places such as restaurants or cafes. Access to the news through TV is therefore different from platforms such as radio or newspapers, as it often involves collective exposure to the broadcasts, and to other viewers' reactions.

---

[41]For the two elections, the estimates are very close to Table 3.3 and the difference between the two samples is not statistically significant. Therefore, it suggests that it is rather the coverage of recent or ongoing conflict on TV that affects voting choices but that individuals do not react differently to the broadcasts based on their past experiences of political conflict.

Notes: The figure plots the estimates η and κ of model 3.4 and the 95% confidence intervals. The circle coefficients are the average effect on the total sample. The diamonds (resp. squares) are the effects among pro-incumbent (resp. pro-incumbent) polling stations. The treatment "Direct" stands for exposure to protests only, "Indirect", exposure through TV only, and "Direct × Indirect" exposure to TV and protests. The reference group is polling stations without TV where no violence occurred.

Figure 4.3: Effect of direct or indirect exposure to TV on turnout

To provide more evidence of this social dimension associated with TV consumption, I use household consumption data from the "Kenyan Integrated Budget Survey" conducted in 2016 by the *Kenyan National Bureau of Statistics* and study whether gaining access to TV changes people's likelihood of attending public places. More precisely, I merge the survey answers with the data on TV access, and restrict the sample to households living in a ward with access to TV in 2017, in order to test if the year in which the signal was introduced correlates with consumption in restaurants, hotels or cafes. I estimate the following model:

$$restaurant_{iwr} = \alpha + \beta.TV15wr + \phi.Urban_{iwr} + \lambda.X_{iwr} + \Gamma_r + u_{iwr} \tag{3.5}$$

where i is a household living in a ward w that was covered by the signal in 2017, in county r. $restaurant_{iwr}$ is a binary variable equal to 1 if one member of the household has consumed food or beverages in a restaurant or an in the past 7 days, or the logarithm of the amount spent in restaurants or hotels in the past 7 days. $TV15_{wr}$ is equal to 1 if the ward gained access to the signal before or in 2015.[42]. I include a dichotomous variable standing for the rural, urban or peri-urban status or the enumeration area of the household, which I interact with the treatment in an alternative specification. X_{iwr} is a set of household covariates including access to electricity, access to the Internet, type of dwelling and availability hand-washing facilities. I add county fixed-effects and cluster the standard errors at the ward level.

Results of model 3.5 are shown in Appendix Table C.23. In columns 1 and 2, I use frequenting of restaurants or hotels as outcomes. I find that households that already had access to TV in 2015, i.e. at the time of the survey, visited these places significantly more often (+4pp). The estimates displayed in column 2 suggest that the effect is mainly driven by rural places.[43] Regarding the amount spent in eateries, I also find that it positively correlates with access to TV in 2015: households who had access to a signal at the time of the survey spent on average 16% more than households living in regions that did not have yet access. This finding may stem from households spending more time in restaurants, or from higher prices in establishments that are likely to be equipped with a TV set. While the data does not allow me to distinguish between these two possible explanations, these findings suggest a differentiated pattern in the frequenting of public places. They are consistent with the following narrative: the increased focus on the electoral crisis in the news broadcasts, which are often watched by voters in public places, may have resulted in social coordination

---

[42]The survey took place at the end of 2015 and beginning of 2016, then $TV15_{wr}$ stands for whether the household had already access to the signal at the time of the interview or if it enters the reach area of a transmitter later (in 2016 and 2017)

[43]The magnitude of the estimates interacted with the urbanization status are similar in magnitude, around 4pp, but the standard deviation of the coefficient is smaller in rural areas.

mechanisms among pro-incumbent supporters, which would explain the positive effect of TV-watching among these voters. Conversely, collective exposure in places that favored the opposition – whose leader was the first to contest the results of the first polls and call for the resignation of the electoral commission – could have exacerbated the perception of uncertainty and deterred people from taking part in to the second ballot.

5   Conclusion

In recent years, access to information in developing countries has undergone a series of profound transformations that are redefining invidual attitudes towards the news:: the enhanced access to technology, increasing media liberalization of the media, and the emergence of social networks. When these changes occur alongside political shocks or episodes of crisis, the influence of media exposure – both in terms of availability of news sources and actual coverage by journalists – can affect the public's beliefs and political behavior.

In this paper, I investigate the impact of TV on turnout and compares how it varies between periods of stability and crisis. To do so, I use the nullification of the 2017 presidential election in Kenya by the Supreme Court and the organization of a new ballot three months later. I estimate a difference-in-differences model and look at the two elections separately. First, I find that exposure to the broadcasts did not influence turnout in the first *regular* election. However, it significantly impacted turnout in the repeat election and this effect is heterogeneous across the political camps. While turnout decreases in pro-opposition polling stations when they are covered by a signal, participation is amplified among regions that voted predominantly for the incumbent President in the previous polls.

To uncover the mechanisms that lie behind this heterogeneity result, I first focus on the content of the news broadcast on TV. Using text analysis methods, I find that before the repeat election, TV channels place a strong emphasis on the ongoing crisis and I provide evidence of a drop in the diversity of news topics covered on TV. This focus on the crisis may have increased risk aversion among the opponent's supporters while encouraging the pro-government side to signal its support for the leader. Moreover, I also find that while the distance to the nearest protest also influenced turnout, the effect is offset by exposure to TV, suggesting that TV is a very powerful medium in periods of political instability. The strength of this medium can be explained by collective exposure to TV in public places. I provide evidence that frequenting of cafes or restaurants is higher in places covered by TV, signal and suggest that this collective dimension of news watching can translate into voting coordination mechanisms.

This article contributes to the growing debate on the transformations taking place in

people's access to information in developing countries. It suggests that individual reactions to uncertainty can be amplified by the media, notably through shifts in editorial choices, which can result in a change in political behavior.

References

Alesina, A., Michalopoulos, S., and Papaioannou, E. (2016). Ethnic inequality. *Journal of Political Economy*, 124(2):428–488.

Angelucci, C. and Cagé, J. (2019). Newspapers in Times of Low Advertising Revenues. *American Economic Journal: Microeconomics.*

Armand, A., Atwell, P., and Gomes, J. F. (2020). The reach of radio: Ending civil conflict through rebel demobilization. *American Economic Review*, 110(5):1395–1429.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly Journal of Economics*, 131(4):1593–1636.

Banerjee, A. V., Kumar, S., Pande, R., and Su, F. (2011). Do Informed Voters Make Better Choices? Experimental Evidence from Urban India.

Bauer, M., Blattman, C., Chytilová, J., Henrich, J., Miguel, E., and Mitts, T. (2016). Can war foster cooperation? *Journal of Economic Perspectives*, 30(3):249–74.

Bellows, J. and Miguel, E. (2009). War and local collective action in sierra leone. *Journal of public Economics*, 93(11-12):1144–1157.

Berger, D., Easterly, W., Nunn, N., and Satyanath, S. (2010). Commercial Imperialism? Political Influence and Trade During the Cold War. NBER Working Papers 15981, National Bureau of Economic Research, Inc.

Besley, T. and Burgess, R. (2002). The Political Economy of Government Responsiveness: Theory and Evidence from India. *The Quarterly Journal of Economics*, 117(4):1415–1451.

Blattman, C. (2009). From Violence to Voting: War and Political Participation in Uganda. *American Political Science Review*, 103(2):231–247.

Branch, D. (2011). *Kenya: Between Hope and Despair, 1963-2010*. Yale University Press.

Brown, R., Montalva, V., Thomas, D., and Velásquez, A. (2019). Impact of violent crime on risk aversion: Evidence from the mexican drug war. *Review of Economics and Statistics*, 101(5):892–904.

Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., and Padro i Miquel, G. (2015). The Value of Democracy: Evidence from Road-Building in Kenya. *American Economic Review*, 105(6):1817—-1851.

Cagé, J. (2015). The Economics of the African Media. In Monga, C. and Lin, J. Y., editors, *The Oxford Handbook of Africa and Economics: Policies and Practices*, pages 605–625. Oxford University Press.

Cagé, J. and Godechot, O. (2017). Who owns the media? The Media Independence Project.

Callen, M., Isaqzadeh, M., Long, J. D., and Sprenger, C. (2014). Violence and risk preference: Experimental evidence from afghanistan. *American Economic Review*, 104(1):123–48.

Casey, K. (2015). Crossing party lines: The effects of information on redistributive politics. *American Economic Review*, 105(8):2410–2448.

Chong, A., De La O, A. L., Karlan, D., and Wantchekon, L. (2015). Does corruption information inspire the fight or quash the hope? a field experiment in mexico on voter turnout, choice, and party identification. *The Journal of Politics*, 77(1):55–71.

Chong, A., León-Ciliotta, G., Roza, V., Valdivia, M., and Vega, G. (2019). Urbanization patterns, information diffusion, and female voting in rural paraguay. *American Journal of Political Science*, 63(2):323–341.

Collier, P. and Vicente, P. C. (2014). Votes and violence: Evidence from a field experiment in Nigeria. *Economic Journal*, 124(574):327–356.

Della Vigna, S., Enikolopov, R., Mironova, V., Petrova, M., and Zhuravskaya, E. (2014). Cross-border media and nationalism: Evidence from serbian radio in croatia. *American Economic Journal: Applied Economics*, 6(3):103–132.

DellaVigna, S. and La Ferrara, E. (2015). Economic and Social Impacts of the Media. pages 723–768.

Di Tella, R. and Franceschelli, I. (2011). Government Advertising and Media Coverage of Corruption Scandals. *American Economic Journal: Applied Economics*, 3(4):119–151.

Dupas, P. and Robinson, J. (2012). The (hidden) costs of political instability: Evidence from kenya's 2007 election crisis. *Journal of Development Economics*, 99(2):314–329.

Durante, R. and Knight, B. (2012). Partisan control, media bias, and viewer responses: Evidence from berlusconi's italy. *Journal of the European Economic Association*, 10(3):451–481.

Durante, R., Pinotti, P., and Tesei, A. (2019). The political legacy of entertainment tv. *American Economic Review*, 109(7):2497–2530.

Easterly, W. and Levine, R. (1997). Africa's growth tragedy: policies and ethnic divisions. *The Quarterly Journal of Economics*, 112(4):1203–1250.

Ellingsen, S. and Hernæs, Ø. (2018). The impact of commercial television on turnout and public policy: Evidence from Norwegian local politics. *Journal of Public Economics*, 159.

Enikolopov, R., Petrova, M., and Zhuravskaya, E. (2011). Media and Political Persuasion: Evidence from Russia. *American Economic Review*, 101(7):3253–85.

Falck, O., Heblich, S., and Gold, R. (2014). E-lections: Voting Behavior and the Internet. *American Economic Review*, 104(7):2238–2265.

Ferraz, C. and Finan, F. (2008). Exposing corrupt politicians: the effects of brazil's publicly released audits on electoral outcomes. *The Quarterly Journal of Economics*, 123(2):703–745.

Freedom House (2016). Kenya — Freedom House. https://freedomhouse.org/report/freedom-press/2017/kenya.

Fujiwara, T. and Wantchekon, L. (2013). Can informed public deliberation overcome clientelism? Experimental evidence from Benin. *American Economic Journal: Applied Economics*, 5(4):241–255.

Gavazza, A., Nardotto, M., and Valletti, T. (2019). Internet and politics: Evidence from uk local elections and local government policies. *The Review of Economic Studies*.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–74.

Gentzkow, M. and Shapiro, J. M. (2006). Media Bias and Reputation. *Journal of Political Economy*, 114(2).

Gentzkow, M., Shapiro, J. M., and Sinkinson, M. (2011). The effect of newspaper entry and exit on electoral politics. *American Economic Review*, 101(7):2980–3018.

Gerber, A. S. and Green, D. P. (2000). The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review*, 94(3):653–663.

Gibson, C. C. and Long, J. D. (2009). The presidential and parliamentary elections in kenya, december 2007. *Electoral studies*, 28(3):497–502.

Guriev, S., Melnikov, N., and Zhuravskaya, E. (2019). 3g internet and confidence in government. *SSRN Working Paper*.

Harwood, A., Herrick, E., and Ugangu, W. (2018). Strengthening Kenyan Media.

Henderson, J., Storeygard, A., and Weil, D. (2009). Measuring economic growth from outer space. *Economic Analysis*, 102(2):994–1028.

Hjort, J. and Poulsen, J. (2019). The arrival of fast internet and employment in africa. *American Economic Review*, 109(3):1032–79.

Horowitz, D. L. (1995). *Ethnic groups in conflict*. University of California Press, Berkeley.

Horowitz, J. and Long, J. (2016). Strategic voting, information, and ethnicity in emerging democracies: Evidence from Kenya. *Electoral Studies*, 44:351–361.

Jensen, R. and Oster, E. (2009). The power of tv: Cable television and women's status in india. *The Quarterly Journal of Economics*, 124(3):1057–1094.

Kasara, K. (2013). Separate and suspicious: Local social and political context and ethnic tolerance in kenya. *the Journal of Politics*, 75(4):921–936.

Knight, B. and Tribin, A. (2019). Opposition media, state censorship, and political accountability: Evidence from chavez's venezuela.

Kramon, E. and Posner, D. N. (2011). Kenya's new constitution. *Journal of democracy*, 22(2):89–103.

La Ferrara, E. (2016). Mass Media and Social Change: Can We Use Television to Fight. *Journal of the European Economic Association*, 14(4):791–827.

Larcinese, V., Puglisi, R., and Snyder, J. M. (2011). Partisan bias in economic news: Evidence on the agenda-setting behavior of U.S. newspapers. *Journal of Public Economics*, 95(9-10):1178–1189.

Lassen, D. D. (2005). The Effect of Information on Voter Turnout: Evidence from a Natural Experiment. *American Journal of Political Science*, 49(1):pp. 103–118.

León, G. (2017). Turnout, political preferences and information: Experimental evidence from peru. *Journal of Development Economics*, 127:56–71.

Long, J. D. and Gibson, C. C. (2015). Evaluating the Roles of Ethnicity and Performance in African Elections: Evidence from an Exit Poll in Kenya. *Political Research Quarterly*, 68(4):830–842.

Marx, B., Pons, V., and Suri, T. (2020). The Perils of Voter Mobilization. *NBER Working Paper*.

Mastrorocco, N. and Minale, L. (2018). News media and crime perceptions: Evidence from a natural experiment. *Journal of Public Economics*, 165:230–255.

McMillan, J. and Zoido, P. (2004). How to Subvert Democracy: Montesinos in Peru. *Journal of Economic Perspectives*, 18(4):69–92.

Michalopoulos, S. and Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary african development. *Econometrica*, 81(1):113–152.

Miguel, E. (2004). Tribe or nation? nation building and public goods in kenya versus tanzania. *World politics*, 56(3):327–362.

Miner, L. (2015). The unintended consequences of internet diffusion: Evidence from Malaysia. *Journal of Public Economics*.

Moya, A. (2018). Violence, psychological trauma, and risk attitudes: Evidence from victims of violence in colombia. *Journal of Development Economics*, 131:15–27.

Nyabola, N. (2018). *Digital Democracy, Analogue Politics: How the Internet Era is Transforming Kenya*. ZED books/African Arguments.

Nyabuga, G. and Booker, N. (2013). Mapping Digital Media: Kenya. Technical report, Open Society Foundations.

Olken, B. A. (2009). Corruption perceptions vs. corruption reality. *Journal of Public Economics*, 93(7-8):950–964.

Pons, V. (2018). Will a five-minute discussion change your mind? A countrywide experiment on voter choice in France. *American Economic Review*, 108(6):1322–1363.

Raleigh, C., Linke, A., Hegre, H., and Karlsen, J. (2010). Introducing acled: an armed conflict location and event dataset: special data feature. *Journal of peace research*, 47(5):651–660.

Ramos, J. et al. (2003). Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, volume 242, pages 133–142. Piscataway, NJ.

Reuter, J. and Zitzewitz, E. (2006). Do ads influence editors? advertising and bias in the financial media. *The Quarterly Journal of Economics*, 121(1):197–227.

Stromberg, D. (2004). Political Mass Media Competition, and Public Policy. *Review of Economic Studies*, 71(1):265–284.

Voors, M. J., Nillesen, E. E., Verwimp, P., Bulte, E. H., Lensink, R., and Van Soest, D. P. (2012). Violent conflict and behavior: a field experiment in burundi. *American Economic Review*, 102(2):941–64.

Wairuri, K. (2017). Ridng the Crest of the Wave? The 2017 election and stagnation of Kenya's democratization process. *Note de l'Observatoire de l'Afrique de l'Est*.

Yanagizawa-Drott, D. (2014). Propaganda and Conflict: Theory and Evidence From the Rwandan Genocide. *Quarterly Journal of Economics*, 129(4):1947–1994.

Appendices

A Content analysis

This Appendix provides details about the analysis of TV headlines in Section 5. After presenting the data and the cleaning procedure, I detail how I construct the indicators of conflict-related stories, of the predictive score of the corpus' words at the monthly level and the concentration of coverage index.

A.1 Preparation of the text dataset

The dataset collected on the website KenyaMoja consists of 421,796 headlines of videos aired between 2013 and 2018 on 5 major Kenyan TV stations (KBC, NTV, Citizen TV, K24, KTN, see Appendix Figure B.8). The following paragraph presents the data cleaning procedure applied on the 2017 data subset (i.e. the sample used for the discussion of the mechanisms in section 4.1).
The 2017 dataset includes 86,189 headlines, summary statistics are shown in Appendix Table C.4.

- **Preliminary cleaning:** After collecting the data, the first step of the cleaning consists in removing the punctuation, figures and stopwords from the dataset. I use the list of stopwords from the Gensim Python package.
- **Names, places and bigrams:** To identify the names of places, I aggregate different layers of administrative data to create lists of local references. For names of individuals, I start with existing lists of public figures to replace the different references to a unique person with a specific token: for instance, the references to the Kenya President "Uhuru Kenyatta", "Kenyatta", "President Uhuru" would return a token of the form "Firstname+Surname", so here "UhuruKenyatta". While such an automatic cleaning is straightforward for sufficiently known figures, names of Kenyan public figures or individual that can be mentioned in a specific news story need to be addressed in a similar way. Therefore, for each fortnight of data, I extract the top 200 tokens to look for possible names or surnames and replace them by the token "Firstname+Surname" after manually checking that were cited in a news story. This procedure does not allow me to deal with all names in the dataset but rather to create a database of individuals whose appearances in the news can be followed in my dataset. One of the issues of text analysis at the word level, as stressed by Ramos et al. (2003), is that it doesn't automatically consider bigrams or groups of words, that have a proper sense only if

252

considered together. In my setting, this is particularly relevant for expressions of interest such as "Supreme Court". To address this type of concern, I repeat the exercise done with the names and identify the words within the top 200 words in a fortnight that are to be categorized as bigrams and fill a dictionary of expressions or bag of words to replace in the headlines.

- **Translation from Swahili to English:** The second step is to homogenize the language of the headlines. The majority of the news items are in English, though some titles were totally written in Swahili ($\sim 5\%$ of the headlines) and more importantly, many English headlines included words in Swahili. For example, the word "maadamano", that can be translated as "protest", was often used in English sentences to refer to the post-electoral protests. To the best of my knowledge, there does not exist translation Python libraries or tools that allows to translate texts from Swahili to English with a sufficiently and meaningfully high success rate. To overcome this limitation and to do the best use of computing power with respect to the mix between the two languages in many headlines, I apply the following procedure. I create a corpus of words that appear more than 10 times in the headlines and implement a language detection algorithm to distinguish between words in English and in Swahili. For the words in Swahili, I use the library Python PyDictionary and create a dictionary. To deal with the words in Swahili that were not categorized as such or were not translated by the algorithm, I manually add them with their translation in the dictionary for the words whose frequency was big enough to bias the results if they had remained untranslated with their English counterpart also appearing in the corpus. Then, I use the dictionary to translate the words in Swahili in the headlines.

A.2 Conflict related stories

In the first part of paragraph 4.1, I analyze the share of conflict-related stories out of the set of headlines for a given day. To do so, I construct a corpus of words that contains words that belong to the lexicon of political conflict or instability. The first set of words does not necessarily relate to violent conflict as it incorporates words that rather characterize political disagreement. I also construct a more restricted set of keywords that more directly relate to the lexicon violence. The words included in the two corpus are shown in Table A.1. For each headline, I construct a dichotomous variable that takes the value 1 if at least one word belonging to one of the sets is cited. Then, I aggregate the data at the daily level for each channel to estimate the share of stories that contain at least one of the keyword, which I define as the conflict-related share of stories in the analysis (see Figure 4.1 and Appendix Figures B.11 and B.12).

A.3 Tf-idf

Setting Given the heterogeneity in the number of news stories aired on each channel (from 4,083 on KBC to 33,390 on NTV), I conduct the analysis with either all channels considered, or analyzing each channel separately before aggregating the final results to get

mean indicators. In the two approaches, I define as a document a month of the year.[44] After the preliminary cleaning presented in paragraph A.1, the second step is to lemmatize the corpus of words – that is to say to retrieve the root of the word so that variations in tense, form, or plural for instance are considered as the same token in the final analysis. I use the Python library Gensim to tokenize and lemmatize the data. Then, I proceed to a first computation of *tf-idf* scores as presented in section 4.1, removing tokens that appear in more than 33% of headlines or less than 1%. These thresholds are standard in the literature and the results are barely affected when moving the lower or upper bounds to 5 or 50%.

Then, I renew the previous cleaning procedure but as a preliminary step, I correct the initial dataset based on the latter dictionary before the lemmatization of the data. The final corpus consists of 26,980 tokens and is used to compute the final *tf-idf* scores.

Tf-idf scores I compute the *inverse document frequency (idf)*, where D is the number of documents in the corpus (here the months of a given year or group of years) containing a token i, and df_i is the number of documents containing token i:

$$idf_i = ln(\frac{D}{df_i})$$

I obtain the *tf-idf* score by multiplying the *inverse document frequency* of token i idf_i by its frequency in document d $\in \{1, ..., D\}$, $tf_{d,i}$:

$$tfidf_i = idf_i \times tf_{d,i}$$

By construction, the score is larger for tokens that are frequent in one document but less in others. The higher the score, the more relevant that token is in a particular document. In our setting, it allows to retrieve the words that are the most prevalent during a given month, in relative terms vis-a-vis the other months, and to infer the most topical news items at different points of time.

Concentration of coverage index The last indicator I construct is the concentration of coverage index. It aims at providing a proxy of the diversity of the news stories during a given period, based on the *tf-idf* scores previously computed. I construct a dataset that includes the Z tokens with the largest scores – where Z is a fixed threshold – and compute the index by summing the squares of the *tf-idf* scores of the top words used during a given month. For each month $m \in \mathbb{M} = \{1, \ldots M\}$, I compute the following index $score_{it}$, which is the score of word i in month t:

$$concentration_m = \sum_{\substack{i \in \{1,...,N\}, \\ t \in \mathbb{M}}} (score_{it})^2 . \mathbb{1}\{t = m\}$$

A greater index $concentration_m$ implies a more concentrated coverage, as the square of the words with the greatest scores (i.e. the most predictive words) has an increasing effect on

---

[44]Following Gentzkow et al. (2019), the definition of a document should be (or related to) the final level of observation relevant to the research question).

the concentration index. Conversely, if in a certain month, the Z most prevalent tokens have lower *tf-idf* scores, the concentration index will be smaller, implying a more general or more diverse news coverage over the period. In other words, a higher index for a particular month tells us that the set of the Z most predictive words have a greater predictive power compared to other months and that the probability of appearance of a subset of these Z tokens is greater. In the analysis of Section 4.1, I set $Z \in \{50, 250\}$ to focus on the most predictive topics but the index can be computed with larger values of Z.
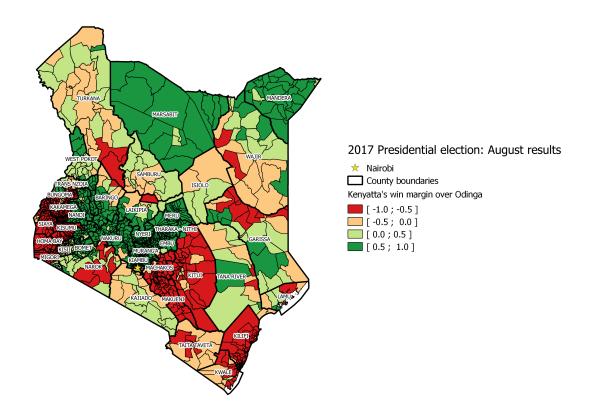
Table A.1: Corpus of conflict-related tokens

(a) Conflict related lexicon

Corpus a
anti-IEBC
assault
attack
barricade
battle
brutality
casualty
clash
commission
conflict
contest
crisis
death
demonstration
discord
disorder
dispute
disrupt
emergency
fight
injure
instability
kill
mobilization
oppose
police
pressure
protest
riot (anti-riot)
strike
struggle
teargas
turmoil
uncertainty
upheaval
violence
war
wound

(b) Violence related lexicon

Corpus b
assault
attack
barricade
battle
brutal
casualty
clash
conflict
contest
crisis
demonstration
disorder
dispute
disrupt
disturb
injure
kill
police
protest
riot (anti-riot)
striker
teargas
turmoil
upheaval
violence
war
wound

B   Additional Figures to Chapter 3

Notes: The map displays the county averages score gap between the two candidates Uhuru Kenyatta and Raila Odinga in the first presidential election (August 8th 2017).
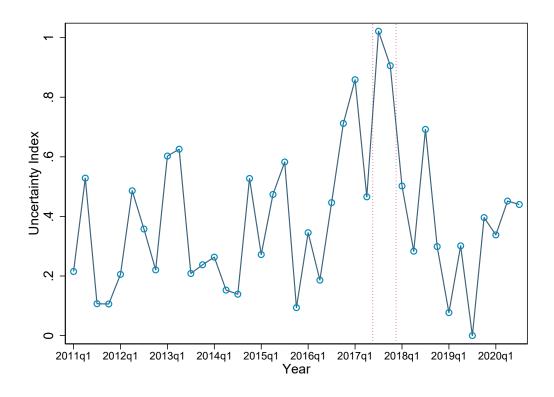
Figure B.1: Margin of the first ranked candidate in August 2017

Notes: The headlines are from the Daily Nation, published on October 4th 2017 and October 10th.

Figure B.2: The 2017 electoral crisis: Newspaper headlines

Notes: The figure plots the uncertainty index in Kenya at the quarter level between 2011 and 2020. Source:
Word Uncertainty Index (`https://www.policyuncertainty.com/`)

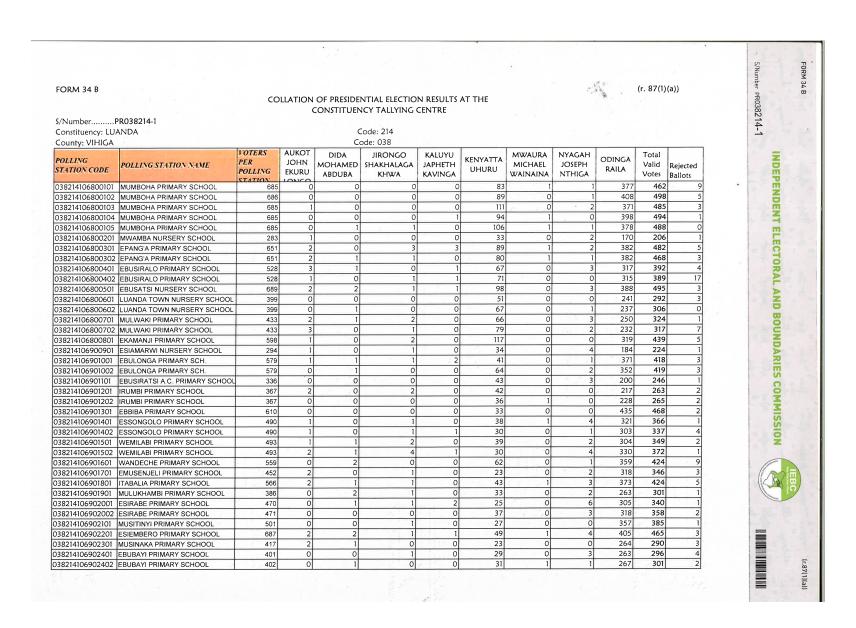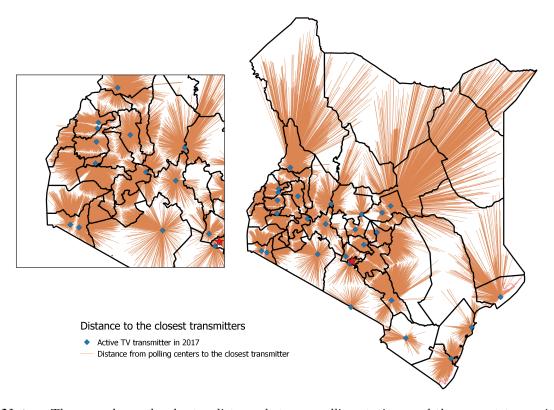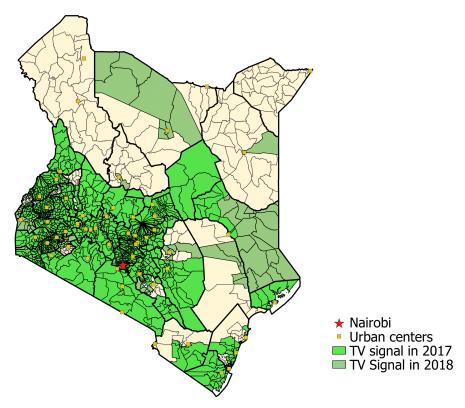Figure B.3: World Uncertainty Index in Kenya (2011-2020)

FORM 34 B

COLLATION OF PRESIDENTIAL ELECTION RESULTS AT THE
CONSTITUENCY TALLYING CENTRE

(r. 87(1)(a))

S/Number..........PR038214-1
Constituency: LUANDA   Code: 214
County: VIHIGA   Code: 038

POLLING STATION CODE	POLLING STATION NAME	VOTERS PER POLLING STATION	AUKOT JOHN EKURU LONGO	DIDA MOHAMED ABDUBA	JIRONGO SHAKHALAGA KHWA	KALUYU JAPHETH KAVINGA	KENYATTA UHURU	MWAURA MICHAEL WAINAINA	NYAGAH JOSEPH NTHIGA	ODINGA RAILA	Total Valid Votes	Rejected Ballots
038214106800101	MUMBOHA PRIMARY SCHOOL	685	0	0	0	0	83	1	1	377	462	9
038214106800102	MUMBOHA PRIMARY SCHOOL	686	0	0	0	0	89	0	1	408	498	5
038214106800103	MUMBOHA PRIMARY SCHOOL	685	1	0	0	0	111	0	2	371	485	3
038214106800104	MUMBOHA PRIMARY SCHOOL	685	0	0	0	1	94	1	0	398	494	1
038214106800105	MUMBOHA PRIMARY SCHOOL	685	0	1	1	0	106	1	1	378	488	0
038214106800201	MWAMBA NURSERY SCHOOL	283	1	0	0	0	33	0	2	170	206	1
038214106800301	EPANG'A PRIMARY SCHOOL	651	2	0	3	3	89	1	2	382	482	5
038214106800302	EPANG'A PRIMARY SCHOOL	651	2	1	1	0	80	1	1	382	468	3
038214106800401	EBUSIRALO PRIMARY SCHOOL	528	3	1	0	1	67	0	3	317	392	4
038214106800402	EBUSIRALO PRIMARY SCHOOL	528	1	0	1	1	71	0	0	315	389	17
038214106800501	EBUSATSI NURSERY SCHOOL	689	2	2	1	1	98	0	3	388	495	3
038214106800601	LUANDA TOWN NURSERY SCHOOL	399	0	0	0	0	51	0	0	241	292	3
038214106800602	LUANDA TOWN NURSERY SCHOOL	399	0	1	0	0	67	0	1	237	306	0
038214106800701	MULWAKI PRIMARY SCHOOL	433	2	1	2	0	66	0	3	250	324	1
038214106800702	MULWAKI PRIMARY SCHOOL	433	3	0	1	0	79	0	2	232	317	7
038214106800801	EKAMANJI PRIMARY SCHOOL	598	1	0	2	0	117	0	0	319	439	5
038214106900901	ESIAMARWI NURSERY SCHOOL	294	1	0	1	0	34	0	4	184	224	1
038214106901001	EBULONGA PRIMARY SCH.	579	1	1	1	2	41	0	1	371	418	3
038214106901002	EBULONGA PRIMARY SCH.	579	0	1	0	0	64	0	2	352	419	3
038214106901101	EBUSIRATSI A.C. PRIMARY SCHOOL	336	0	0	0	0	43	0	3	200	246	1
038214106901201	IRUMBI PRIMARY SCHOOL	367	2	0	2	0	42	0	0	217	263	2
038214106901202	IRUMBI PRIMARY SCHOOL	367	0	0	0	0	36	1	0	228	265	2
038214106901301	EBBIBA PRIMARY SCHOOL	610	0	0	0	0	33	0	0	435	468	2
038214106901401	ESSONGOLO PRIMARY SCHOOL	490	1	0	1	0	38	1	4	321	366	1
038214106901402	ESSONGOLO PRIMARY SCHOOL	490	1	0	1	1	30	0	1	303	337	4
038214106901501	WEMILABI PRIMARY SCHOOL	493	1	1	2	0	39	0	2	304	349	2
038214106901502	WEMILABI PRIMARY SCHOOL	493	2	1	4	1	30	0	4	330	372	1
038214106901601	WANDECHE PRIMARY SCHOOL	559	0	2	0	0	62	0	1	359	424	9
038214106901701	EMUSENJELI PRIMARY SCHOOL	452	2	0	1	0	23	0	2	318	346	3
038214106901801	ITABALIA PRIMARY SCHOOL	566	2	1	1	0	43	1	3	373	424	5
038214106901901	MULUKHAMBI PRIMARY SCHOOL	386	0	2	1	0	33	0	2	263	301	1
038214106902001	ESIRABE PRIMARY SCHOOL	470	0	0	1	2	25	0	6	305	340	1
038214106902002	ESIRABE PRIMARY SCHOOL	471	0	0	0	0	37	0	3	318	358	2
038214106902101	MUSITINYI PRIMARY SCHOOL	501	0	0	1	0	27	0	0	357	385	1
038214106902201	ESIEMBERO PRIMARY SCHOOL	687	2	2	1	1	49	1	4	405	465	3
038214106902301	MUSINAKA PRIMARY SCHOOL	417	2	1	0	0	23	0	0	264	290	3
038214106902401	EBUBAYI PRIMARY SCHOOL	401	0	0	1	0	29	0	3	263	296	4
038214106902402	EBUBAYI PRIMARY SCHOOL	402	0	1	0	0	31	1	1	267	301	2

S/Number PR038214-1

FORM 34 B

INDEPENDENT ELECTORAL AND BOUNDARIES COMMISSION

IEBC

[r.87(1)(a)]

Figure B.4: Example of a form 34B (Luanda Constituency in Vihiga County)

Distance to the closest transmitters

◆ Active TV transmitter in 2017

— Distance from polling centers to the closest transmitter

Notes: The map shows the shorter distance between polling stations and the nearest transmitter.

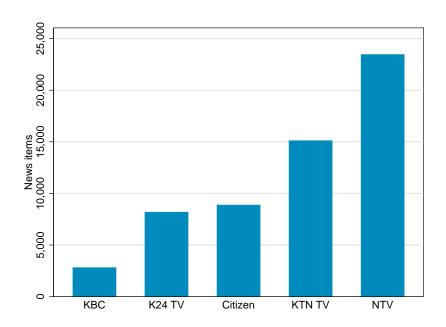Figure B.5: Distance from polling centers to nearest transmitter

Notes: The map shows the ward with no access to TV in 2017 or that got access to TV before 2018 or in the year 2018.
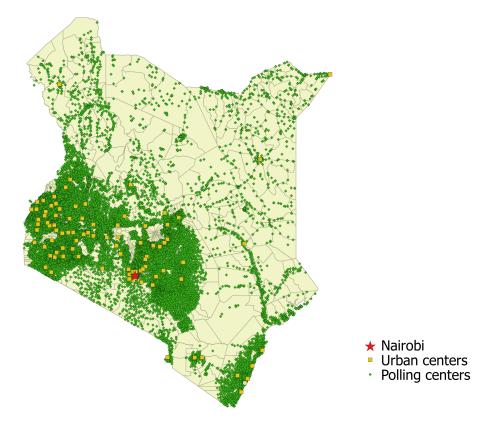
Figure B.6: Access to TV in 2018

Notes: The figure shows an example a news story broadcast by KBC on November 10th, 2020, scrapped from the website KenyaMoja.

Figure B.7: Example of a news story
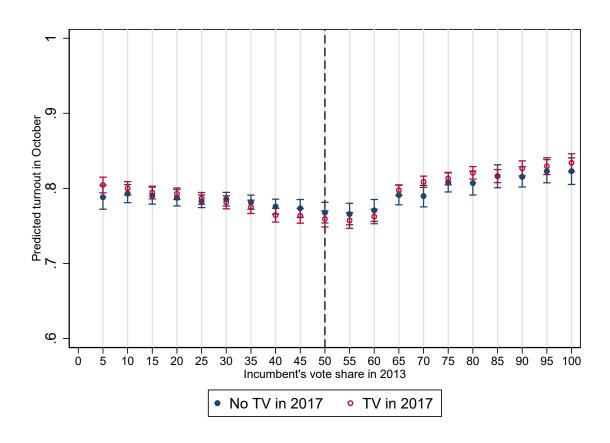
Notes: The figure shows the number of headlines collected for each channel in 2017. Source: KenyaMoja.

Figure B.8: News items by channel
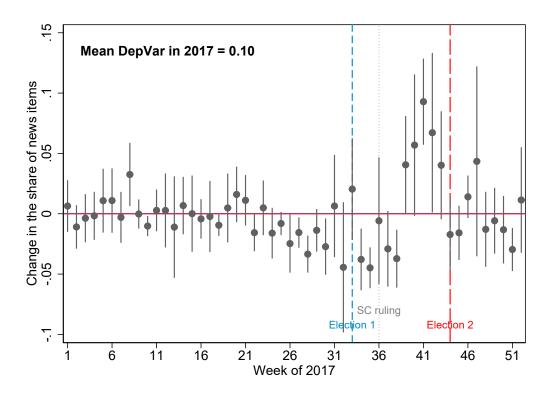
Notes: The map shows the geo-location of Kenyan polling centers and of the main urban centers.

Figure B.9: Polling centers and main urban centers in Kenya
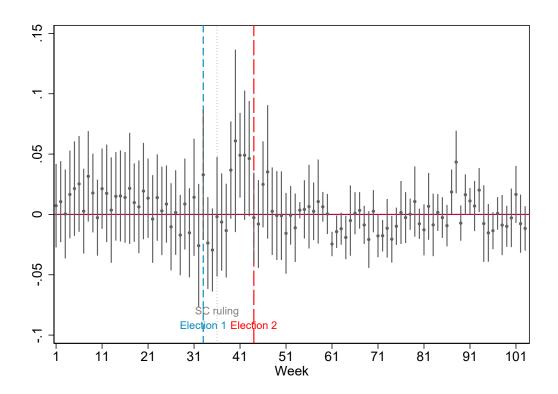
Notes: The figure displays the predicted turnout in August 2017 in polling stations covered by the TV signal vs. not covered. Turnout is predicted for each quantile of Kenyatta's score in the first election: for instance, the coefficients at x=100 are estimated on polling stations where the incumbent score in August was above the 95th threshold. The reference category is the first quantile (5% of polling stations where Kenyatta got his lowest score in August).

Figure B.10: Marginal impact of TV on turnout in the repeated election of October
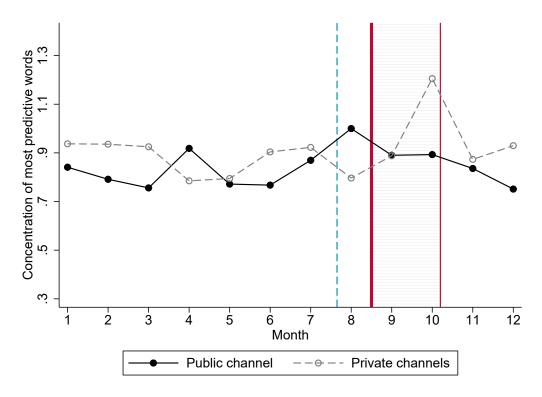
Notes: The figure plots the estimates retrieved from the regression of the share of conflict related stories per day on week fixed effect. The share of conflict related stories is computed using the restricted set of keywords (see Appendix A.1b). The reference category is the average share of conflict related stories over the year i.e. 0.8. The green vertical line stands for the week of the first election (August 8th 2017), the red line for the week of the Supreme Court decision to nullify the election (September 1st 2017) and the blue line for the recall election (October 26th 2017).

Figure B.11: Share of conflict and electoral related content

Notes: The figure plots the estimates retrieved from the regression of the share of conflict related stories per day on week fixed effect in 2017 and 2018. The share of conflict related stories is computed using the restricted set of keywords (see Appendix A.1a). The reference category is the average share of conflict related stories over the year i.e. 0.8. The green vertical line stands for the week of the first election (August 8th 2017), the red line for the week of the Supreme Court decision to nullify the election (September 1st 2017) and the blue line for the recall election (October 26th 2017).

Figure B.12: Share of conflict-related content

Notes: The figure plot the concentration index as defined in subsection 4.1 and compare the four most popular private channels in Kenya (Citizen TV, NTV, KTN, K24), with the state-owned channel KBC.

Figure B.13: News concentration - State-owned vs private channel

C Additional Tables to Chapter 3

Table C.1: Summary statistics: Violence in Kenya between 2007 and 2013

	Mean	Sd	Min	Max	Sum	N
At least one violence episode:						
Year before the election	0.18	0.38	0	1	259	1,450
Between 2013 and 2017	0.33	0.47	0	1	482	1,450
Between 2007 and 2013	0.28	0.19	0	1	399	1,450
2007 Post electoral violence	0.11	0.99	0	1	162	1,450
Year before the election	3.36	6.07	1	75	870	259
Share of riots	0.81	0.40	0	1	-	259
Between 2013 and 2017	5.34	13.74	1	230	2573	482
Share of riots	0.72	0.45	0	1	-	482
Between 2007 and 2013	5.30	14.40	1	222	2116	399
Share of riots	0.67	0.47	0	1	-	399
2007 post electoral violence	3.51	6.50	1	47	569	162
Share of riots	0.80	0.40	0	1	-	162

Notes: An observation is a ward. Source: ACLED. Reading: Out of the 1450 wards, 259 (18%) experienced at least one violent episode in the year before the election. In this period, the average number of events was 3.36 and the maximum number of recorded events in a ward is 870.

Table C.2: Device Ownership in Kenya

	R3 2005-2006	R4 2008-2009	R5 2010-2011	R6 2014-2015	R7 2016-2017
Own a radio set (%)	79.6	85.7	80.9	80.6	85.4
Own a TV set (%)	25.5	32.1	37.8	40.1	42.5
Own a computer (%)	-	-	-	-	10.7

Notes: An observation is an household. Source: Afrobarometer - Rounds 3 to 7 Survey question: Which of these things do you personally own?

Table C.3: TV Access at the Sublocation Level

	Mean	sd	Min	Max	N
Analog TV in 2013	0.44	0.50	0	1	7,167
TV in 2013 (Analog+Digital)	0.46	0.50	0	1	7,167
TV in 2015 (Analog+Digital)	0.53	0.50	0	1	7,167
Digital TV in 2016	0.60	0.49	0	1	7,167
Digital TV in 2017	0.74	0.44	0	1	7,167
Digital TV in 2018	0.84	0.37	0	1	7,167

Notes: An observation is a sublocation. Source: Communcation Authority of Kenya. Before 2013, both analog and digital signal were available: the analog switch-off (ASO) occured in 2015. After the ASO - i.e. from 2016 onwards - the only type of available signal is digital.

Table C.4: Summary statistics: News stories in 2017

	Mean	sd	Min	Max	N
News items	235.90	58.89	6	398	365
Conflict related items	22.02	10.57	2	82	365
Share of conflict related items	0.15	0.06	0.02	0.39	365

Notes: The table displays summary statistics on news stories broadcast on TV in 2017. An observation is a day. The five Kenyan main channels are included: KBC, Citizen TV, NTV, K24, and KTN TV.

Table C.5: Robustness Checks: Predictors of TV signal availability

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					TV in 2017					
Distance transmitter (km)	-0.369**	-0.315*	-0.598***	-0.369***	-0.469***	-0.315***	-0.530***	-0.565***	-0.375**	-0.510***
	(0.170)	(0.159)	(0.082)	(0.090)	(0.102)	(0.097)	(0.118)	(0.152)	(0.146)	(0.154)
Mean altitude (m)	0.024	0.138	-0.279***	0.024	-0.134	0.138*	-0.237**	0.045	0.042	0.182*
	(0.131)	(0.106)	(0.083)	(0.081)	(0.103)	(0.083)	(0.104)	(0.101)	(0.140)	(0.104)
Std altitude (m)	-0.009	-0.009	-0.005	-0.009	-0.002	-0.009	-0.003	-0.009	-0.000	-0.009
	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)	(0.007)	(0.010)	(0.008)	(0.010)	(0.008)
Mean altitude ward	0.002	0.025	0.071*	0.002	0.096***	0.025	-0.018	0.002	0.011	0.026
	(0.036)	(0.030)	(0.037)	(0.037)	(0.034)	(0.038)	(0.050)	(0.051)	(0.049)	(0.052)
Std altitude ward	-0.016	-0.025	-0.017	-0.016	-0.021	-0.025	0.021	-0.015	0.021	-0.025
	(0.021)	(0.017)	(0.016)	(0.018)	(0.016)	(0.019)	(0.024)	(0.024)	(0.024)	(0.024)
Distance to capital (km)		-0.118			0.159	-0.118			0.689**	0.173
		(0.213)			(0.189)	(0.162)			(0.350)	(0.358)
Distance to the coast (km)		-0.219			0.413	-0.219			0.333	-0.447
		(0.853)			(0.421)	(0.366)			(0.447)	(0.381)
Dist to closest town (km)		0.008			-0.186	0.008			-0.270**	-0.015
		(0.165)			(0.118)	(0.101)			(0.135)	(0.122)
Observations	19,502	19,502	19,502	19,502	19,502	19,502	11,358	11,358	11,358	11,358
Mean DepVar	0.70	0.70	0.70	0.70	0.70	0.70	0.56	0.56	0.56	0.56
Sd DepVar	0.46	0.46	0.46	0.46	0.46	0.46	0.50	0.50	0.50	0.50
Adjusted R2	0.59	0.59	0.37	0.59	0.40	0.59	0.33	0.50	0.36	0.50
Cluster	Transmitter	Transmitter	Ward	Ward	Ward	Ward	Transmitter	Transmitter	Transmitter	Transmitter
Transmitter FE			✓		✓		✓		✓	
County FE	✓	✓		✓		✓		✓		✓
Transmitter polynomial	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Topography	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Location controls	✓	✓	✓	✓	✓	✓			✓	✓
TV in 2013	✓	✓	✓	✓	✓	✓				

Notes: Standard errors in parentheses are clustered at the closest transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is a polling center. Transmitter polynomial include the fourth polynomial of the distance between the polling center and the closest transmitter. Topography includes altitude and the standard deviation of altitude in the Voronoi polygon of the center, as well as similar statistics computed at the ward level. Location controls include distance to the capital, to the coast, to the closest town above 30,000 inhabitants, to the closest urban center, as well as the squared distances. Pop-Eco controls include population in the ward, density, light at night in the ward (mean and standard deviation). County fixed effect are included in columns 1, 2 4, 6, 8 and 10. Standard errors are clustered at the closest transmitter level in in columns 1, 2, and 7 to 10, and at the ward level in columns 3 to 6. In columns 9 and 10, polling stations with access to analog TV in 2013 are excluded.

Table C.6: Correlation between TV access and socio-economic covariates

	DHS 2014			Household survey 2016					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Work for pay	Informal	1ry educ.	2ry education	Hh size	Food	Consumption	Hand-washing	Internet
TV in 2017	2.908*	-0.025	9.494**	4.513***	-0.887***	929.481***	2,672.342***	-0.110***	-0.150***
	(1.544)	(0.495)	(3.767)	(1.580)	(0.191)	(189.810)	(519.128)	(0.025)	(0.030)
Observations	18,022	18,022	18,022	18,022	1,055	1,055	1,055	1,047	1,055
Mean DepVar	17.70	11.99	54.29	18.27	4.02	4270.91	8087.71	1.78	1.69
Sd DepVar	10.34	6.00	13.93	9.31	1.22	2033.38	5094.60	0.27	0.30
Adjusted R2	0.12	0.01	0.12	0.30	0.08	0.03	0.04	0.02	0.04
County FE									
Transmitter FE									
Propagation controls									

Notes Standard errors in parentheses are clustered at the transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. TV in 2017 is a binary variable equal to 1 if TV is available in the sublocation. An observation is a polling station (columns 1 to 4) or an household in 2016. TV in 2017: binary variable equal to 1 if TV is available in the sublocation. Work for pay, Informal, 1ry and 2ndy education are measured as the share of the population in the ward. Food and Consumption are in KES. Hand-washing/Internet are equal to 1 if the household has no access, 2 otherwise. Model estimated without propagation controls.

Table C.7: Correlation TV access and political and community covariates

	2013 election			Community			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Turnout in 2013	Incumbent score	Opposition score	Share 1st ethnic group	Incumbent coethnic	Opposition coethnic	Violence 07-13
TV in 2017	0.017***	-0.019	0.004	-0.407***	-0.052	0.084**	-0.065*
	(0.005)	(0.029)	(0.028)	(0.110)	(0.033)	(0.037)	(0.036)
Observations	18,022	18,022	18,022	18,022	18,022	18,022	18,022
Mean DepVar	0.88	0.47	0.46	3.00	0.32	0.37	0.26
Sd DepVar	0.11	0.41	0.39	1.41	0.47	0.48	0.44
Adjusted R2	0.01	0.08	0.11	0.05	0.13	0.07	0.00
County FE							
Transmitter FE							
Propagation controls							

Notes Standard errors in parentheses are clustered at the transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is a polling station. TV in 2017 is a binary variable equal to 1 if TV is available in the sublocation. Share first ethnic group is the quantile of the share of the dominant ethnic group in the ward as a proxy for the degree of ethnic homogeneity. Incumbent (resp Opposition) co-ethnic is a binary variable equal to 1 if the dominant group in the ward is part of Kenyatta's (resp Odinga) political coalitions. Violence 2007-2013 is equal to one when at least one episode of violence is recorded in the ward by ACLED across the period. The model is estimated without propagation controls.

Table C.8: Exogeneity checks: socio-economic covariates

	DHS 2014			Household survey 2016					
	(1) Work for pay	(2) Informal	(3) 1ry educ.	(4) 2ry education	(5) Hh size	(6) Food	(7) Consumption	(8) Hand-washing	(9) Internet
TV in 2017	-0.718	-0.203	-0.934	-0.067	-0.053	156.713	295.341	-0.016	0.001
	(0.878)	(0.869)	(0.768)	(0.396)	(0.171)	(267.468)	(366.943)	(0.016)	(0.017)
Observations	18,022	18,022	18,022	18,022	861	861	861	854	861
Mean DepVar	17.70	11.99	54.29	18.27	3.97	4371.93	8399.96	1.77	1.67
Sd DepVar	10.34	6.00	13.93	9.31	1.23	2131.37	5286.00	0.28	0.30
Adjusted R2	0.64	0.27	0.88	0.85	0.53	0.30	0.55	0.36	0.55
County FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Transmitter FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Propagation controls	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes Standard errors in parentheses are clustered at the transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. TV in 2017 is a binary variable equal to 1 if TV is available in the sublocation. An observation is a polling station (columns 1 to 4) or an household in 2016. TV in 2017: binary variable equal to 1 if TV is available in the sublocation. Work for pay, Informal, 1ry and 2ndy education are measured as the share of the population in the ward. Food and Consumption are in KES. Hand-washing/Internet are equal to 1 if the household has no access, 2 otherwise.

Table C.9: Exogeneity checks: 2008 socio-economic covariates

	DHS 2008						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Hh size	Electricity	Radio	Poor	Medium	Rich	Years of educ
TV in 2017	-0.105	0.019	0.006	-0.031	0.039	0.052	-0.108
	(0.084)	(0.027)	(0.044)	(0.027)	(0.040)	(0.055)	(0.083)
Observations	32,185	32,185	32,185	32,185	32,185	32,185	22,970
Mean DepVar	4.86	0.15	0.73	0.21	0.21	0.20	4.26
Sd DepVar	1.43	0.36	0.44	0.41	0.41	0.40	2.48
Adjusted R2	0.13	0.48	0.13	0.11	0.10	0.11	0.01
County FE	✓	✓	✓	✓	✓	✓	✓
Transmitter FE	✓	✓	✓	✓	✓	✓	✓
Propagation controls	✓	✓	✓	✓	✓	✓	✓

Notes Standard errors in parentheses are clustered at the transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is an household in 2008 surveyed for the Demographic and Health Survey. TV in 2017: binary variable equal to 1 if TV is available in the sublocation. Electricity, Radio, Poor, Medium, Rich are dichotomous variables.

Table C.10: Exogeneity checks: political and community covariates, restricted sample

	2013 election			Community			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Turnout in 2013	Incumbent score	Opposition score	Share 1st ethnic group	Incumbent coethnic	Opposition coethnic	Violence 2007-2013
TV in 2017	-0.006	0.015	-0.026	0.157	-0.002	-0.013	-0.006
	(0.005)	(0.016)	(0.019)	(0.210)	(0.005)	(0.027)	(0.037)
Observations	10,544	10,544	10,544	10,544	10,544	10,544	10,544
Mean DepVar	0.87	0.37	0.57	3.23	0.18	0.27	0.27
Sd DepVar	0.13	0.39	0.38	1.35	0.38	0.44	0.44
Adjusted R2	0.25	0.80	0.78	0.64	0.91	0.94	0.27
County FE	✓	✓	✓	✓	✓	✓	✓
Transmitter FE	✓	✓	✓	✓	✓	✓	✓
Propagation controls	✓	✓	✓	✓	✓	✓	✓

Notes Standard errors in parentheses are clustered at the transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is a polling station without access to TV in 2013. TV in 2017 is a binary variable equal to 1 if TV is available in the sublocation. Share first ethnic group is the quantile of the share of the dominant ethnic group in the ward as a proxy for the degree of ethnic homogeneity. Incumbent (resp Opposition) co-ethnic is a binary variable equal to 1 if the dominant group in the ward is part of Kenyatta's (resp Odinga) political coalitions. Violence 2007-2013 is equal to one when at least one episode of violence is recorded in the ward by ACLED across the period.

Table C.11: Exogeneity checks: political and community covariates, restricted sample

	DHS 2014			Household survey 2016					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Work for pay	Informal	1ry educ.	2ry education	Hh size	Food	Consumption	Hand-washing	Internet
TV in 2017	-0.728	0.026	-1.235*	0.303	-0.001	-47.513	-103.905	-0.004	0.024
	(0.823)	(1.087)	(0.635)	(0.495)	(0.226)	(273.236)	(392.233)	(0.018)	(0.020)
Observations	10,544	10,544	10,544	10,544	456	456	456	449	456
Mean DepVar	14.90	12.48	52.14	14.34	4.58	3724.90	6082.19	1.85	1.83
Sd DepVar	8.62	7.10	16.59	7.50	1.26	2225.68	3682.26	0.21	0.22
Adjusted R2	0.61	0.25	0.93	0.85	0.43	0.06	0.48	0.49	0.61
County FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Transmitter FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Propagation controls	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes Standard errors in parentheses are clustered at the transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is a polling station (columns 1 to 4) or an household in 2016 without access to TV in 2013. TV in 2017: binary variable equal to 1 if TV is available in the sublocation. Work for pay, Informal, 1ry and 2ndy education are measured as the share of the population in the ward. Food and Consumption are in KES. Hand-washing/Internet are equal to 1 if the household has no access, 2 otherwise.

Table C.12: Access to TV and turnout in open polling stations

	Turnout in August			Turnout in October		
	(1)	(2)	(3)	(4)	(5)	(6)
TV in 2017	0.003	-0.001	0.006	-0.009	-0.038**	0.022
	(0.005)	(0.004)	(0.006)	(0.007)	(0.015)	(0.018)
Incumbent vote Sh. in 2013	0.033**		0.023*	0.443***		0.355***
	(0.013)		(0.013)	(0.061)		(0.064)
Pro Incumbent 2013		0.009			0.168***	
		(0.006)			(0.035)	
TV 2017 × Pro Incumbent 2013		0.011			0.102***	
		(0.008)			(0.026)	
TV 2017 × Incumbent Vote Sh.			0.016			0.150***
			(0.013)			(0.043)
Observations	15,946	15,946	15,946	15,946	15,946	15,946
Mean DepVar	0.79	0.79	0.79	0.41	0.41	0.41
Sd DepVar	0.09	0.09	0.09	0.31	0.31	0.31
Adjusted R2	0.51	0.51	0.51	0.91	0.89	0.92
County FE	✓	✓	✓	✓	✓	✓

Notes: Standard errors in parentheses are clustered at the closest transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is polling station that was open for the two rounds (first and recall election). Pro Incumbent 2013 is a binary variable equal to 1 if the incumbent Kenyatta ranked first in the 2013 election. Incumbent Vote Sh is continuous and centered around 0. In Columns 1 to 3 (resp. 4 to 6), the dependent variable in turnout in the first election of August (resp. turnout in the recall election of October).

Table C.13: Access to TV and turnout - alternative clustering

	Turnout in August			Turnout in October		
	(1)	(2)	(3)	(4)	(5)	(6)
TV in 2017	-0.000	-0.002	0.003	-0.005	-0.034***	0.029***
	(0.003)	(0.003)	(0.004)	(0.008)	(0.008)	(0.009)
Incumbent vote Sh. in 2013	0.024***		0.017**			0.339***
	(0.006)		(0.007)			(0.025)
Pro Incumbent 2013		0.007			0.156***	
		(0.005)			(0.017)	
TV 2017 × Pro Incumbent 2013		0.009			0.111***	
		(0.005)			(0.017)	
TV 2017 × Incumbent Vote Sh. 2013			0.011			0.161***
			(0.007)			(0.021)
Observations	18,000	18,000	18,000	15,946	15,946	15,946
Mean DepVar	0.79	0.79	0.79	0.41	0.41	0.41
Sd DepVar	0.09	0.09	0.09	0.31	0.31	0.31
Adjusted R2	0.52	0.52	0.52	0.86	0.90	0.92
County FE	✓	✓	✓	✓	✓	✓

Notes: Standard errors in parentheses are clustered at the ward level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is polling station that was open for the two rounds (first and recall election). Pro Incumbent 2013 is a binary variable equal to 1 if the incumbent Kenyatta ranked first in the 2013 election. Incumbent Vote Sh is continuous and centered around 0. In Columns 1 to 3 (resp. 4 to 6), the dependent variable in turnout in the first election of August (resp. turnout in the recall election of October).

Table C.14: First-difference approach

	Turnout in October		
	(1)	(2)	(3)
TV in 2017	-0.015	-0.032*	0.001
	(0.009)	(0.016)	(0.014)
Incumbent vote Sh. in August	0.570***		0.513***
	(0.049)		(0.047)
Pro Incumbent August		0.203***	
		(0.026)	
TV 2017 × Pro Incumbent August		0.058**	
		(0.026)	
TV 2017 × Incumbent Vote Sh. in Aug			0.098***
			(0.024)
Observations	15,944	15,946	15,944
Mean DepVar	0.41	0.41	0.41
Sd DepVar	0.31	0.31	0.31
Adjusted R2	0.94	0.91	0.95
County FE	✓	✓	✓

Notes: Standard errors in parentheses are clustered at the closest transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is polling station open in the recall election. Pro Incumbent 2013 is a binary variable equal to 1 if the incumbent Kenyatta ranked first in the first election of August. Incumbent Vote Sh is continuous and centered around 0.

Table C.15: Access to TV and turnout - additional controls

	Turnout in August			Turnout in October		
	(1)	(2)	(3)	(4)	(5)	(6)
TV in 2017	0.000	-0.002	0.004	-0.010	-0.028*	0.014
	(0.004)	(0.004)	(0.006)	(0.008)	(0.014)	(0.014)
Incumbent vote Sh. in 2013	0.033**		0.025	0.326***		0.268***
	(0.015)		(0.015)	(0.058)		(0.057)
Pro Incumbent 2013		0.008			0.099***	
		(0.006)			(0.024)	
TV 2017 × Pro Incumbent 2013		0.011			0.063***	
		(0.007)			(0.018)	
TV 2017 × Incumbent Vote Sh. 2013			0.016			0.114***
			(0.011)			(0.030)
Observations	18,007	18,011	18,007	15,953	15,957	15,953
Mean DepVar	0.79	0.79	0.79	0.41	0.41	0.41
Sd DepVar	0.09	0.09	0.09	0.31	0.31	0.31
Adjusted R2	0.52	0.52	0.52	0.92	0.91	0.92
County FE	✓	✓	✓	✓	✓	✓
TV × Signal controls	✓	✓	✓	✓	✓	✓

Notes: Standard errors in parentheses are clustered at the closest transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is polling station where the two rounds took place (first and recall election). Pro Incumbent 2013 is a binary variable equal to 1 if the incumbent Kenyatta ranked first in 2013. Incumbent Vote Sh is continuous and centered around 0. We add the interaction between the propagation controls with the pro-incumbent measure to equation 3.3.

Table C.16: Placebo test: access to TV in 2018

	Turnout in August			Turnout in October		
	(1)	(2)	(3)	(4)	(5)	(6)
TV in 2018	0.004	0.005	0.004	0.002	-0.001	0.003
	(0.005)	(0.005)	(0.005)	(0.010)	(0.016)	(0.010)
Incumbent vote Sh. in 2013	0.015*		0.018**	0.429***		0.428***
	(0.008)		(0.007)	(0.030)		(0.033)
Pro Incumbent		0.006**			0.168***	
		(0.003)			(0.019)	
TV 2018 × Pro Incumbent		-0.002			0.001	
		(0.005)			(0.018)	
TV 2018 × Incumbent Vote Sh. 2013			-0.006			0.004
			(0.006)			(0.021)
Observations	5,663	5,663	5,663	5,168	5,168	5,168
Mean DepVar	0.78	0.78	0.78	0.33	0.33	0.33
Sd DepVar	0.09	0.09	0.09	0.27	0.27	0.27
Adjusted R2	0.45	0.45	0.45	0.88	0.85	0.88
County FE	✓	✓	✓	✓	✓	✓

Notes: Standard errors in parentheses are clustered at the closest transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is polling station that wasn't covered by the signal in 2018. Pro Incumbent 2013 is a binary variable equal to 1 if the incumbent Kenyatta ranked first in 2013. Incumbent Vote Sh is continuous and centered around 0. In Columns 1 to 3 (resp. 4 to 6), the dependent variable in turnout in the first election of August (resp. turnout in the recall election of October).

Table C.17: Robustness check: later access to TV

	Turnout in August			Turnout in October		
	(1)	(2)	(3)	(4)	(5)	(6)
TV in 2017	-0.012	-0.013*	-0.014	-0.018	-0.043**	0.026
	(0.007)	(0.007)	(0.012)	(0.015)	(0.021)	(0.021)
Incumbent vote Sh.	0.019		0.021**	0.290***		0.252***
	(0.012)		(0.009)	(0.051)		(0.044)
Pro Incumbent		0.010**			0.123***	
		(0.004)			(0.023)	
TV 2017 × Pro Incumbent		0.007			0.101***	
		(0.016)			(0.031)	
TV 2017 × Incumbent Vote Sh.			-0.006			0.172***
			(0.026)			(0.031)
Observations	7,265	7,265	7,265	5,903	5,903	5,903
Mean DepVar	0.79	0.79	0.79	0.32	0.32	0.32
Sd DepVar	0.09	0.09	0.09	0.26	0.26	0.26
Adjusted R2	0.48	0.48	0.48	0.82	0.81	0.83
County FE	✓	✓	✓	✓	✓	✓

Notes: Standard errors in parentheses are clustered at the closest transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is polling station without access to TV in 2015. Pro Incumbent 2013 is a binary variable equal to 1 if the incumbent Kenyatta ranked first in 2013. Incumbent Vote Sh is continuous and centered around 0. In Columns 1 to 3 (resp. 4 to 6), the dependent variable in turnout in the first election of August (resp. turnout in the recall election of October).

Table C.18: Most prevalent tokens in 2017 by month

rank	January	February	March
1	registr	registr	unair
2	CORD (Coalition for Reforms and Democracy)	unair	tribut
3	email	CHAP (Community Health Access Program)	old
4	punchlin	lofti	babi
5	healthcrisis	healthcrisi	healthcrisis
6	snapshot	punchlin	bearer
7	Gambia	CORD (Coalition for Reforms and Democracy)	IOC (International Olympic Committee)
8	UDM (United Democratic Movement)	sportpesa	sack
9	BarackObama	progress	Somali
10	unair	allstar	joint
11	goe	KMPDU	InternationalWomensDay
12	outbreak	ball	semi
13	AFCON	valentin	CHAP (Community Health Access Program)
14	YayaJammeh	hull	hire
15	mistari	intervent	think
16	KMPDU (Kenya Medical Practitioners Union)	blackstreet	fatten
17	diary	wed	lawless
18	mode	red	lift
19	DonaldTrump	financi	read
20	storm	hammock	agent

	April	May	June
1	bearer	endors	manifesto
2	flag	MothersDay	muslim
3	princip	summit	special
4	lump	labour	drank
5	talli	gymnasium	figur
6	bra	tribun	palasso
7	armi	expand	deed
8	thousand	verif	civil
9	easter	Dubai	grab
10	ensur	havoc	choral
11	parallel	distribut	clean
12	amid	backdrop	print
13	conclud	ride	MRFC (Mwamba Rugby Football Club)
14	muti	kaligraph	abdidor
15	worm	unmarri	IDP (Internally Displaced Persons)
16	contract	recruit	mustard
17	retreat	detect	corpor
18	pari	Mexico	hunter
19	deliveri	jackpot	HIV
20	provison	standard	ramadhan

rank	July	August	September
1	tribut	server	TheMaragaDecision
2	heat	AfriCOG (African Center for Open Governance)	judgement
3	debate	role	KECOSO (Kenya Communications Sports Organization)
4	frontlin	accept	memo
5	eulog	newli	cabinet
6	moth	cast	deleg
7	interior	big	heartfelt
8	print	KHRC (Kenya Human Rights Commission)	surgeon
9	endors	JohnKerry	dissent
10	hurdl	oath	cyber
11	striker	grant	lioness
12	latin	prejudic	grandstand
13	popat	HongKong	berlin
14	discu	githeri	storm
15	manifesto	swear	affair
16	credibl	petition	milgren
17	bisil	sita	lancet
18	KTDA (Kenya Tea Development Agency)	usurp	chan
19	request	myriad	cab
20	jam	mountain	political

	October	November	December
1	anti_iebc	KCPE (Kenya Certificate of Primary Education)	christmas
2	MashujaaDay	exam	edit
3	edit	societi	swear
4	malaria	RobertMugabe	enteract
5	dialogu	inauguration	boda
6	commission	execut	kiran
7	gospel	civil	CECAFA (Council for East and Central Africa Football Asso)
8	anti_riot	entertain	holiday
9	toll	secess	daili
10	includ	midsiz	cricket
11	put	verdict	graft
12	FIFA	diseas	perish
13	lift	jfk	boiler
14	riot	dialogu	Dubai
15	particip	Maasai	agenda
16	prepar	multipurpos	left
17	feed	chach	newsmak
18	insist	crew	earli
19	obsolet	forward	valu
20	panafrican	waterfal	biro

Notes: The table presents the top 20 words with the highest tf idf score for each month of 2017. Names of local places and individuals are removed from the list. Channels included: KBC, Citizen TV, KTN, NTV, K24 TV.

Table C.19: Direct and indirect exposure to conflict

	Turnout in October	
	(1)	(2)
Treatment		
Violence	-0.004	0.013
	(0.014)	(0.018)
TV	-0.012	-0.039**
	(0.008)	(0.014)
TV × Violence	-0.023**	-0.042**
	(0.009)	(0.015)
Treatment × Pro Incumbent		
Violence × Pro Incumbent		-0.020
		(0.033)
TV × Pro Incumbent		0.103***
		(0.027)
Violence × TV × Pro Incumbent		0.090***
		(0.026)
Observations	15,972	15,972
Mean DepVar	0.41	0.41
Sd DepVar	0.31	0.31
Adjusted R2	0.91	0.91
County FE	✓	✓
$\beta i = \beta j$		
Violence=TV	.609	.019
Violence=TV x Violence	.275	.017
TV=TV x Violence	.074	.660
x Pro Incumbent		
Violence =TV		0.00
Violence =TV x Violence		.001
TV=TV x Violence		.203

Notes: Standard errors in parentheses are clustered at the closest transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is polling station opened in October 2017. Violence equals 1 when at least one episode of protests occurred in the ward between the two rounds. TV equals 1 when the polling center got access to TV between 2013 and 2017. In column 2, the treatment variable is interacted with a dichotomous variable equal to 1 if the Incumbent ranked first in the previous election of 2013. $\beta i = \beta j$ reports the p-value of the χ^2 tests of coefficient equality.

Table C.20: Distance to the closest protests and exposure to TV

| | Turnout in October 2017 | | | |
	(1)	(2)	(3)	(4)
TV	-0.011	-0.025***	-0.011	-0.028**
	(0.008)	(0.008)	(0.008)	(0.012)
Distance closest violence	0.001	0.006***	0.000	0.005*
	(0.001)	(0.001)	(0.001)	(0.002)
TV × distance	-0.000	-0.004***	-0.001	-0.004*
	(0.001)	(0.001)	(0.001)	(0.002)
× Pro Incumbent				
TV × pro Incumbent		0.068***		0.078***
		(0.016)		(0.023)
Distance closest violence		-0.012***		-0.009***
		(0.002)		(0.003)
TV × distance		0.010***		0.007**
		(0.002)		(0.003)
Model	Distance closest betw. 2 rounds		Distance closest in 2017	
Observations	15,972	15,972	15,972	15,972
Mean DepVar	0.41	0.41	0.41	0.41
Sd DepVar	0.31	0.31	0.31	0.31
Adjusted R2	0.91	0.91	0.91	0.91
County FE	✓	✓	✓	✓

Notes: Standard errors in parentheses are clustered at the closest transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is polling station opened in October 2017. TV equals 1 when the polling center got access to TV between 2013 and 2017. Distance to the closest episode of violence is continuous and measured in 100 meter units. In columns 2 and 4, the treatment variable is interacted with a dichotomous variable equal to 1 if the Incumbent ranked first in the previous election of 2013.

Table C.21: Distance to the closest protests and exposure to TV - (crisis hotspots excluded)

| | Turnout in October 2017 | | | |
	(1)	(2)	(3)	(4)
TV in 2017	-0.011	-0.023***	-0.011	-0.027**
	(0.008)	(0.008)	(0.008)	(0.012)
Distance closest violence	0.001	0.006***	0.000	0.005*
	(0.001)	(0.001)	(0.001)	(0.002)
TV × distance	-0.000	-0.004***	-0.001	-0.004*
	(0.001)	(0.001)	(0.001)	(0.002)
× Pro Incumbent				
TV × pro Incumbent		0.066***		0.077***
		(0.016)		(0.023)
Distance closest violence		-0.012***		-0.009***
		(0.002)		(0.003)
TV × distance		0.010***		0.007**
		(0.002)		(0.003)
Model	Distance closest betw. 2 rounds		Distance closest in 2017	
Observations	15,672	15,672	15,672	15,672
Mean DepVar	0.41	0.41	0.41	0.41
Sd DepVar	0.31	0.31	0.31	0.31
Adjusted R2	0.91	0.91	0.91	0.91
County FE	✓	✓	✓	✓

Notes: Standard errors in parentheses are clustered at the closest transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is polling station opened in October 2017. TV equals 1 when the polling center got access to TV between 2013 and 2017. Distance to the closest episode of violence is continuous and measured in 100 meter units. In columns 2 and 4, the treatment variable is interacted with a dichotomous variable equal to 1 if the Incumbent ranked first in the previous election of 2013. The three largest cities, where the density of protests was higher, are excluded from the sample (Nairobi, Kisumu, Mombasa).

Table C.22: Long-run exposure to violence and turnout

	Turnout in August				Turnout in October			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Protests between 2013 and 2017					
	No	Yes	No	Yes	No	Yes	No	Yes
TV	-0.004	0.006	-0.006	0.003	-0.012	0.009	-0.035**	-0.027
	(0.005)	(0.006)	(0.005)	(0.005)	(0.009)	(0.018)	(0.014)	(0.024)
Pro Incumbent 2013	0.017*	0.011*	0.012	0.006	0.240***	0.195***	0.182***	0.135***
	(0.009)	(0.006)	(0.008)	(0.008)	(0.043)	(0.048)	(0.042)	(0.044)
TV × Pro Incumbent in 2013			0.008	0.009			0.096**	0.119***
			(0.011)	(0.008)			(0.035)	(0.029)
Observations	12,041	5,988	12,041	5,988	10,618	5,354	10,618	5,354
Mean DepVar	0.79	0.80	0.79	0.80	0.40	0.43	0.40	0.43
Sd DepVar	0.08	0.09	0.08	0.09	0.31	0.30	0.31	0.30
Adjusted R2	0.45	0.54	0.45	0.54	0.90	0.86	0.91	0.87
County FE	✓	✓	✓	✓	✓	✓	✓	✓
TV_noViolence=TV.Violence	ref	.073	ref	.095	ref	.276	ref	.681
TV_noViolence x Pro Inc.=TV.Violence x Pro Inc			ref	.966			ref	.6

Notes: Standard errors in parentheses are clustered at the closest transmitter level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is polling station one the two rounds took place. TV equals when the polling center got access to TV between 2013 and 2017. In columns 3, 4, the treatment variable is interacted with a dichotomous variable equal to 1 if the Incumbent ranked first in the previous election of 2013. In columns 1 to 4 (resp. 5 to 8), the dependent variable is turnout in the August 2017 election (resp. October). The sample is split into polling stations for which at least one political protest (resp. no protest) occurred in the ward between 2013 and 2017 in even columns (resp. uneven columns).

Table C.23: Frequenting of TV and consumption in restaurants

	Attended		Spent	
	(1)	(2)	(3)	(4)
Got TV in 2015	0.039***		0.162***	
	(0.014)		(0.053)	
TV in 2015 × Rural		0.038***		0.159***
		(0.012)		(0.055)
TV in 2015 × Urban		0.043		0.160
		(0.041)		(0.204)
TV in 2015 × Peri-Urban		0.042		0.195
		(0.044)		(0.217)
Observations	14,757	14,757	14,757	14,757
Mean DepVar	0.19	0.19	1.00	1.00
Sd DepVar	0.40	0.40	2.21	2.21
Adjusted R2	0.10	0.10	0.11	0.11
County FE	✓	✓	✓	✓
Households controls	✓	✓	✓	✓

Notes: Standard errors in parentheses are clustered at the ward level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. An observation is an household surveyed in 2015 or 2016 for the *Kenyan Integrated Household Budget Survey*, who lives in a ward where TV signal was available in 2017. The dependent variable *Attended* is a binary variable equal to 1 if one member of the household has attended a restaurant, hotel or cafe in the past 7 days, *Spent* is equal to the log+1 of the amount spent in catering places. In columns 1 and 3, the treatment is a dichotomous variable equal to 1 if the household had access to TV before the survey. In columns 2 and 4, I interact the treatment with urbanization status (rural, peri-urban or urban).

Résumé

Trois essais en économie politique de
l'information

Introduction

Cette thèse porte principalement sur l'étude des déterminants de l'offre du marché de l'information, et cherche à examiner comment ces multiples facteurs peuvent influencer nos sociétés et nos systèmes politiques. Les trois chapitres sont indépendants les uns des autres et peuvent être lus séparément. Ils constituent trois études dans le champ de l'économie des médias et visent à contribuer au débat sur les modes de production de l'information et leurs possibles effets.

Pour mettre en évidence les relations entre un déterminant donné de la fonction de production de l'information et du discours politique, j'utilise une variété de méthodes, tenant principalement de la microéconomie appliquée. J'emprunte également aux méthodes utilisées en science politique et en sciences de la communication, afin de mieux mettre en perspective et interpréter les résultats tirés de l'analyse quantitative des données.

Dans le premier chapitre, nous analysons les paramètres qui affectent la capture (ou captation) de médias, et comment les perceptions de la probabilité qu'une enquête journalistique aboutisse à la publication d'une information inédite, ainsi que les caractéristiques du marché influencent les différentes formes de pression exercées sur les médias dans le monde.

Dans le deuxième chapitre, nous étudions l'influence de l'argent sur le discours des hommes et femmes politiques : il est question de comprendre comment le fait de recevoir des dons de la part d'entreprises privées peut affecter le discours des candidats. Pour ce faire, nous analysons les professions de foi publiées par les candidats aux élections législatives françaises dans les années 1990.

Dans le troisième chapitre, je m'intéresse aux possibles impacts du changement technologique sur la production des contenus et le comportement politique des citoyens. Plus précisément, j'étudie l'effet de l'introduction de la télévision numérique au Kenya sur l'information et comment les programmes ont affecté les préférences politiques lors de l'élection présidentielle de 2017.

Ainsi, les trois chapitres de cette thèse visent à illustrer la complexité du paysage médiatique et des déterminants du discours politique. En m'appuyant sur les résultats existants de la littérature et en utilisant un large éventail de méthodes, cette thèse cherche à contribuer au débat général sur les défis relatifs au monde de l'information dans ses multiples aspects.

Chapitre 1

Ce chapitre est co-écrit avec Charles Louis-Sidois.

Le premier chapitre de cette thèse porte sur la capture de médias et ses déterminants. Le phénomène de capture peut être appréhendé à la manière de Stiglitz (2017), qui la définit comme une situation économique où les régulateurs deviennent excessivement favorables aux acteurs qu'ils sont censés réguler. Dans le contexte médiatique, on considérera la capture comme une influence sur le contenu produit par les médias du fait d'influences extérieures visant les producteurs d'information. L'auteur souligne que la capture peut prendre différentes formes, la plus évidente étant par la propriété ou l'actionnariat, mais elle peut se matérialiser sous forme d'incitations financières ou de pressions, telles que la censure. Ces différentes formes ont été largement étudiées dans la littérature. Par exemple, Dyck et al. (2008); Durante and Knight (2012) examinent la capture exercée par les propriétaires de médias ; McMillan and Zoido (2004) se penchent sur les paiements directs faits aux journalistes ; et Gratton (2015) se concentre sur l'impact des menaces directes sur l'activité journalistique. L'objectif de ce chapitre est d'aborder la complexité du phénomène en étudiant conjointement les différents types de capture. Nous avons recours à des méthodes théoriques et empiriques, pour mieux comprendre quels sont les environnements et contextes qui expliquent la prévalence de certains types de pressions à l'encontre des médias.

Différents facteurs contribuent à l'élaboration de l'agenda éditorial d'un média. A ces déterminants s'ajoute la probabilité – telle qu'elle est perçue par les journalistes – qu'une enquête journalistique aboutisse. Ainsi, les coûts prohibitifs que peuvent impliquer une enquête ou la taille limitée de l'audience peuvent dissuader les journalistes de choisir un sujet précis d'investigation. Nous examinons comment ces multiples paramètres entrent dans le choix d'un agent qui aurait intérêt à ce qu'une information ne soit pas révélée au public, et nous introduisons la dimension temporelle de la capture comme un paramètre clé de la stratégie que choisira cet agent.

La distinction principale que nous établissons entre les différentes stratégies de capture de médias est le contrôle de l'agenda journalistique, qui lui, dépend du moment où la capture a lieu. D'un côté du spectre, être propriétaire d'un média permet à un lobby, un homme ou femme politique, ou tout agent désireux d'interférer sur la production d'information, d'obtenir un contrôle total de l'agenda. Cette stratégie est préventive : elle implique une acquisition *ex ante* du média avant qu'aucune enquête ne soit lancée. Nous appelons cette stratégie *capture interne*. A l'inverse, le principal peut s'abstenir de fixer l'ordre du jour et proposer un paiement *ex post* après que le processus d'enquête ait potentiellement abouti. Cela peut par exemple prendre la forme d'un pot-de-vin en échange de la suppression d'une

information possiblement préjudiciable. Nous appelons cette option réactive *capture externe.*

Nous nous appuyons sur les travaux précurseurs de Besley and Prat (2006) et proposons un modèle théorique dans lequel un "principal" (lobby, acteur politique, entreprise, ...) est corrompu. Pour obtenir un signal quant au type du principal, les médias choisissent stratégiquement leur niveau d'investissement dans l'ouverture d'une enquête journalistique. Lorsque qu'un média trouve une information et la révèle au public (qu'il reçoit un signal positif indiquant que le principal est corrompu), celui-ci obtient un gain positif, alors que le principal est négativement affecté. Le principal dispose de deux types de stratégie pour empêcher la divulgation du signal. Premièrement, il peut effectuer un transfert vers un média au début du jeu et l'empêcher d'enquêter (capture interne). Alternativement, le principal peut faire une offre à un média uniquement si ce dernier a choisi d'enquêter et a trouvé un signal positif quant à la corruption du principal, il effectue un transfert en échange de la suppression du signal (capture externe).

Notre modèle théorique permet de mettre en évidence deux principales prédictions. Celles-ci concernent l'effet de la corruption perçue sur la capture des médias. Premièrement, le modèle suggère que la *capture interne* augmente positivement avec la corruption perçue. Un accord précoce est préféré par le principal si les médias déploient des efforts d'investigation soutenus, ce qui est le cas s'ils croient initialement que le principal est corrompu. En ce qui concerne la capture externe, nos résultats mettent en évidence une relation positive avec le niveau de corruption perçue lorsque le marché des médias est compétitif. L'effet est à l'inverse négatif lorsque le marché est moins développé. Cette prédiction peut s'expliquer par deux forces opposées en œuvre : une augmentation de la corruption perçue entraîne davantage d'enquêtes menées par les médias, ce qui augmente la nécessité de recourir à la capture externe. D'autre part, la corruption perçue augmente la capture interne, ce qui rend la capture externe inutile. Lorsque le marché des médias est de taille restreinte, le second effet domine le premier, expliquant ainsi le résultat sur la capture externe.

Dans la deuxième partie du chapitre, nous testons la validité de ces prédictions théoriques sur la base d'une analyse empirique des conditions d'exercice des journalistes dans le monde. Une difficulté méthodologique est de mettre en évidence une relation causale entre choix d'investigation des médias et phénomènes de capture (Prat and Strömberg, 2013). Notre idée ici est d'utiliser un choc exogène sur la probabilité d'initier une enquête journalistique, qui serait indépendant, au moins à court terme, des spécificités nationales des marchés des médias. Pour ce faire, nous utilisons la révélation des *Panama Papers* et l'importante

couverture mondiale dont elle a fait l'objet.[1] Nous exploitons le fait que les pays ont été différemment exposés au choc – le nombre d'entreprises et individus ayant des activités offshore mentionnés dans les *Panama Papers* diffèrent d'un pays à l'autre – et nous examinons ainsi l'effet de ces variations entre pays sur les éventuels phénomènes de capture.

Grâce à une approche empirique dite de "différence de différences", nous montrons tout d'abord que la publication des *Panama Papers* a effectivement eu un effet sur les perceptions de la corruption. En effet, dans les pays les plus exposés au scandale, la corruption perçue a augmenté de 1,2 point de pourcentage à la suite des révélations médiatiques. Nous nous appuyons sur ce résultat pour analyser si ce changement dans les perceptions a affecté le paysage médiatique et la stratégie de capture. Pour ce faire, nous utilisons les données fournies par *Reporters sans Frontières* pour 187 pays entre 2012 à 2018. Pour étudier la capture interne, nous considérons les conflits d'intérêts avec les propriétaires de médias, ainsi que la part des propriétaires ayant des intérêts dans d'autres secteurs économiques. La capture externe est considérée via les paiements externes, comme les pots-de-vin et les avantages en nature accordés aux journalistes.

Nos résultats empiriques sont en accord avec nos prédictions théoriques. Premièrement, nous constatons une augmentation de 9 points de pourcentage de la part des propriétaires ayant des intérêts dans d'autres secteurs économiques, ce qui confirme l'hypothèse théorique d'une relation positive entre perceptions de corruption et capture interne. Deuxièmement, nous mettons en évidence que l'augmentation de la corruption perçue se traduit par une prévalence plus faible de la capture externe dans les pays où le paysage médiatique est le moins développé, alors que nous obtenons le résultat inverse dans les pays où le paysage médiatique est plus compétitif.

Ce chapitre a pour but de mieux comprendre l'étendue de la capture médiatique aujourd'hui et de proposer une nouvelle approche pour aborder sa complexité. La perception de la corruption comme déterminant clé de l'entrée dans les médias est le principal déterminant étudié, mais d'autres paramètres, tels que la concurrence des médias, affectent également le choix du principal. Notre modèle et nos résultats empiriques constituent une première étape vers une analyse plus fine des formes de stratégies de capture, démarche qui nous semble cruciale dans le débat concernant la protection de la liberté de la presse dans le monde.

---

[1]Les *Panama Papers* font référence à une enquête internationale menée sous l'égide du Consortium international des journalistes d'investigation suite à fuite de documents mentionnant explicitement des individus et des entreprises impliqués dans des systèmes d'évasion fiscale. Les publications ont eu lieu en avril 2016 après une année d'enquête.

Chapitre 2

Ce chapitre est co-écrit avec Julia Cagé et Caroline Le Pennec.

Le deuxième chapitre porte sur les effets des dons d'entreprises sur les choix de communication des candidats aux élections en campagne. Plus précisément, nous nous concentrons sur les élections législatives françaises dans les années 1990, et examinons les professions de foi des candidats afin d'étudier l'impact que peuvent avoir les dons de personnes morales sur les choix rhétoriques des candidats. La question de l'argent en politique a été largement étudiée, notamment sous l'angle de la relation entre dépenses de campagne et succès électoral ou activité des élus (voir par exemple Levitt 1994; Ansolabehere et al. 2003; Kalla and Broockman 2018 ou Bekkouche et al. 2020). La façon dont le type de revenus peut affecter directement le discours des candidats a été considérée dans de précédents travaux. Pourtant, nous savons notamment grâce à Feltovich and Giovannoni (2015); Kendall et al. (2015) ou Cruz et al. (2018), que les messages de campagne sont un déterminant important des choix des électeurs.

Pour étudier cette question en détails, nous avons construit une nouvelle base de données qui réunit des informations sur les dons reçus par les candidats se présentant aux élections législatives dans 555 circonscriptions, ainsi les professions de foi qu'ils ou elles ont publiées en amont des élections. Notre base de données contient 10 285 professions de foi, provenant de Le Pennec (2020) et de notre propre travail de collecte aux Archives nationales. De plus, nos données incluent des informations détaillées sur les revenus de campagne, ainsi que les montants donnés par les entreprises à chaque candidat, pour lesquels nous connaissons également les caractéristiques individuelles (parti, sexe, autres mandats politiques, ...).

Nous exploitons un choc historique sur les dons des entreprises et estimons un modèle de "différence de différences" pour isoler l'impact causal des dons sur la communication de campagne des candidats. Un réforme mise en œuvre en 1995 a interdit les dons des entreprises aux candidats politiques : cette loi a marqué un changement important dans le financement des partis politiques et des élections. Seules les personnes "physiques" (c'est-à-dire les individus) sont désormais autorisées à faire des dons aux candidats ou partis politiques.[2]

Le but premier de cette étude est d'évaluer si l'interdiction a eu un effet sur les messages de campagne des candidats, et – si tel était le cas –, quels aspects du discours auraient pu

---

[2]Depuis 1988, la "Commission Nationale des Comptes de Campagne et des Financements Politiques" est chargée de vérifier et d'approuver les comptes des campagnes des candidats. Nous documentons que la réforme de 1995 n'a pas été anticipée par les candidats.

être affectés. Nous formulons quatre hypothèses pour mieux appréhender comment les dons des entreprises pourraient influencer le contenu des professions de foi. Premièrement, les dons d'entreprises pourraient pousser les candidats à mener une campagne centrée sur des enjeux plus locaux – par exemple, axée sur des problématiques propres à leur circonscription – ou plus nationaux - par exemple, suivant le discours des partis et les débats politiques nationaux de premier plan. Deuxièmement, les candidats pourraient avoir recours à un vocabulaire plus polarisé à droite ou à gauche, ou à l'inverse plus modéré en réponse aux dons. Troisièmement, les dons d'entreprises pourraient jouer un rôle dans la prédominance de certains sujets politiques – comme l'économie ou la sécurité – au détriment d'autres, tels que les affaires étrangères ou sociales. Quatrièmement, ils pourraient influencer la qualité de la communication d'un candidat, que nous évaluons en utilisant le niveau d'originalité des professions. Pour faire de ces quatre hypothèses des indicateurs quantitatifs qui peuvent être étudiés économétriquement, nous appliquons diverses méthodes d'analyse de texte – empruntées à Gentzkow et al. (2019) et Le Pennec (2020). Nous construisons quatre mesures distinctes qui représentent les dimensions du langage sur lesquelles les politiciens individuels ont un contrôle et qu'ils peuvent ajuster en réponse aux contributions de campagne qu'ils reçoivent.

Les contributions des entreprises aux campagnes des différents candidats ne sont toutefois pas allouées de façon aléatoire : l'estimation de leur effet causal nécessite de mettre en œuvre une stratégie d'identification qui tient compte de leur nature endogène. En 1993, environ un tiers des candidats aux élections législatives françaises a reçu au moins un don d'une entreprise. En moyenne, les hommes reçoivent plus de dons d'entreprises que les femmes – dont les montants sont également supérieurs – et les candidats sortants et ceux qui se présentent de façon répétée dans leur circonscription reçoivent davantage que les nouveaux candidats. Nous constatons également que les candidats représentant les partis traditionnels reçoivent plus - en nombre de dons et en montant total - que les candidats des partis mineurs ou moins populaires. Nous exploitons le fait qu'avant l'interdiction des dons d'entreprise, les candidats diffèrent de manière substantielle en termes d'argent reçu sous cette forme, et nous estimons un modèle de "différence de différences" pour quantifier l'impact de recevoir de dons d'entreprises en 1993. Nous contrôlons les caractéristiques des candidats qui prédisent les dons, et ajoutons des effets fixes (*candidat* et *année/parti*) afin que les résultats ne soient pas biaisés du fait d'un possible effet de sélection sur les variables inobservables.

Nous estimons ensuite notre modèle empirique en utilisant comme variable dépendante chacun des quatre indices présentés précédemment. Tout d'abord, nous trouvons un impact économiquement significatif des dons d'entreprise sur l'*index local*, ce qui suggère que le

fait de recevoir plus de dons encourage les candidats à davantage mettre en évidence leur présence et action dans la circonscription dans leur profession de foi. D'après nos estimations, une augmentation des dons d'un écart-type augmente l'*index local* de 16% d'écart-type. L'effet est particulièrement élevé pour les partis les moins expérimentés sur la scène politique (candidats verts et d'extrême droite). Deuxièmement, nous ne trouvons aucun impact significatif des dons sur le *score gauche-droite* des professions de foi, ni un quelconque effet de modération du discours. Cependant, en examinant l'hétérogénéité possible entre partis, nous constatons que les candidats plus extrêmes sur l'axe droite-gauche ajustent davantage leur discours. En effet, alors que l'effet sur le *score gauche-droite* reste faible et généralement non significatif pour les trois partis historiquement dominants en France (droite, socialiste et communiste), le vocabulaire des candidats des partis verts se déplacent vers la droite. Les candidats d'extrême-droite eux se déplacent vers la gauche. Troisièmement, nous définissons quatre grandes catégories de sujets de campagne : la sécurité et l'administration, la politique étrangère, l'économie, et les questions sociales. Nous mettons en évidence que les dons d'entreprises encouragent les candidats à davantage aborder des questions économiques dans leurs manifestes, au détriment des sujets de politique étrangère ou relatifs aux questions sociales. Par exemple, une augmentation d'un écart-type des dons d'entreprises augmente la probabilité de traiter principalement de questions économiques de 1,6 point de pourcentage. Enfin, nous documentons un impact significatif et positif des dons sur l'*index d'originalité* : ce résultat suggère que les candidats font plus d'efforts pour rédiger une profession de foi plus personnalisée que les autres candidats du même parti qui reçoivent moins d'argent de la part d'entreprises.

La dernière partie de ce deuxième chapitre vise à discuter et mettre en perspective ces résultats empiriques. Notre interprétation est la suivante : les dons d'entreprises auraient un "effet électoral" sur les candidats. Le fait de recevoir des contributions d'entreprises locales augmenterait l'importance des sujets locaux et économiques dans les campagnes électorales. Lorsque les candidats changent leurs perceptions des sujets qui importent le plus aux citoyens de leur circonscription, ils adapteraient leur stratégie de communication en conséquence. Cet effet est particulièrement significatif pour les petits candidats et les outsiders, dont les convictions préalables quant aux préférences de l'électorat peuvent ne pas être aussi précises que celles des candidats plus implantés.

Cette interprétation est celle que nous préférons, car nous considérons que nos données ne mettent pas en évidence que d'autres mécanismes seraient en œuvre. Par exemple, nous aurions pu considérer que les dons d'entreprises auraient eu un effet positif sur 'l'*index local* dans le discours électoral via raison d' "effet de ressource". Une augmentation des revenus

de campagne, quelle qu'en soit la source, peut permettre aux candidtats de mener une campagne de meilleure qualité, s'appuyant sur une communication plus ciblée, mieux adaptée à l'électorat local. Cependant, une analyse plus détaillée suggère une forte hétérogénéité de l'effet de dons selon la taille du donateur. En effet, nous constatons que l'effet des dons sur l'*index local* est tiré par les plus petits donateurs, c'est-à-dire les entreprises qui ont donné à un seul candidat dans une seule circonscription. Cette hétérogénéité semble exclure un tel mécanisme d'effet de ressource, car ce dernier aurait été pertinent pour tous les types de donateurs. Par ailleurs, d'autres sources de revenus de campagne, telles que les contributions personnelles ou les contributions des partis, ne semblent pas augmenter les références locales de la même manière que les dons d'entreprises.

Une autre explication possible est que les entreprises donatrices peuvent s'attendre à recevoir des avantages particuliers en échange de leurs dons. Cette interprétation de l'effet dit de "quid-pro-quo" est cohérente avec le fait que les donateurs locaux, dont les intérêts sont essentiellement basés dans la circonscription, poussent les candidats à mettre en avant leur présence sur le terrain dans leur manifeste – et à faire valoir l'influence qu'ils peuvent exercer au niveau local pour servir les intérêts de leurs donateurs. Pourtant, nous ne trouvons pas d'impact significatif des dons des entreprises sur la quantité ou le contenu des questions au gouvernement ou dans les interventions des élus au Parlement. Ces résultats ne permettent pas de conclure que les hommes et femmes politiques favorisent leurs donateurs une fois au pouvoir. Certes, nous ne pouvons pas exclure l'existence d'une relation de contrepartie entre les entreprises donatrices et les élus, toutefois il est peu probable qu'un tel mécanisme soit le principal canal dans l'ajustement du discours des candidats.

Dans l'ensemble, nos résultats suggèrent que les liens des candidats avec les entreprises privées peuvent modifier le comportement et le discours des hommes et femmes politiques, même en l'absence d'accords de contreparties directes ou indirectes. Nous montrons que les contributions aux campagnes électorales peuvent influencer les citoyens via leur effet indirect sur les candidats : en affectant la perception qu'ont les candidats des préoccupations de leur électorat, les contributions façonnent le contenu des messages de campagne diffusés dans les professions de foi. En fin de compte, ce sont les informations sur lesquelles s'appuient les électeurs pour prendre leurs décisions de vote qui sont affectées.

Chapitre 3

Dans le troisième chapitre de cette thèse, je me penche sur un exemple de changement technologique dans le paysage médiatique – l'introduction de la télévision numérique au

Kenya – afin d'étudier l'impact que peut avoir le développement d'un nouveau média sur les comportements électoraux. Au Kenya, une augmentation substantielle de la disponibilité du signal télévisé a eu lieu au cours de la dernière décennie : la pénétration graduée de la télévision donne un cadre empirique pertinent pour étudier les retombées de l'entrée d'une nouvelle technologie dans un pays d'Afrique subsaharienne. Entre 2013 et 2017, plus de 2 500 *sublocations* ont bénéficié de la migration de la télévision analogique vers la télévision numérique. Plus de 30% de la population a bénéficié de cette expansion. Cette augmentation de la pénétration de la télévision est en grande partie due à l'installation d'antennes, dans le cadre du plan de migration numérique mené par le gouvernement du Kenya et de l'adoption des normes de transmission *Digital Video Broadcasting - Terrestrial* (DVB-T et DVB-T2), financées et supervisées avec le soutien de l'*Union Internationale des Télécommunications*.

Ce chapitre s'appuie sur la littérature relative à l'impact de la télévision sur les résultats du vote (Gentzkow and Shapiro, 2006; Ellingsen and Hernæs, 2018; Durante et al., 2019) et discute de sa pertinence dans un pays en développement. J'exploite le contexte très particulier de l'élection kényane de 2017 pour examiner plus en profondeur les effets hétérogènes possibles de l'exposition à la télévision, en période de grande incertitude politique. L'annulation de l'élection de 2017 par la Cour suprême a changé la donne politique en l'espace de seulement quelques semaines. En effet, la décision d'annuler l'élection a été annoncée le 1er septembre 2017, juste après que le président sortant ait été déclaré vainqueur du scrutin organisé le 8 août 2017. La plainte déposée par le parti d'opposition dénonçant les incohérences dans la communication des résultats à la commission électorale entre les différents centres de vote, a été jugée recevable par la Cour. Cette déclaration a marqué une rupture dans la politique kényane, et le débat autour de l'annulation et de l'organisation d'un nouveau scrutin ont été accompagnés d'une vague de protestations menée par les leaders de l'opposition (Wairuri, 2017). Le second vote, qui a eu lieu trois mois après la première élection, a été largement boycotté par l'opposition, et son organisation a considérablement augmenté le climat d'incertitude et de crise politique dans le pays. Par conséquent, je compare l'effet de la télévision sur le taux de participation à la première élection - que nous pouvons considérer comme un scrutin ayant eu lieu dans un contexte "normal" - et à la seconde élection. Je discute ainsi la façon dont le climat de crise politique peut être un facteur important dans l'hétérogénéité de l'influence de l'exposition aux médias chez les électeurs.

Avant d'estimer l'impact de la télévision sur la participation électorale, j'étudie les principaux déterminants de la réception de la télévision, sur la base d'une étude de données topographiques et des caractéristiques socio-économiques des ménages. En effet, l'installation d'antennes de télévision et l'exposition aux programmes télévisuels sont très probablement

endogènes. Il est plausible que les antennes aient été installées à des endroits stratégiques, afin d'atteindre le public le plus large possible, et que les nouveaux utilisateurs de télévision aient des attitudes de vote qui ne soient pas représentatives de l'ensemble de la population. Pour mesurer l'influence réelle de l'exposition aux programmes, nous devons formuler l'hypothèse que les zones avec ou sans accès à la télévision en 2017 sont similaires en termes de caractéristiques non observées, conditionnellement aux variables observées. Je construis un ensemble de données géocodées des bureaux de vote et des antennes de télévision au Kenya, et je calcule une fonction polynomiale entre la distance de chaque bureau et l'émetteur le plus proche. La base de données contient également des mesures de diverses caractéristiques topographiques (altitude, terrain plus ou moins accidenté, ...), la distance à la capitale ou à la ville la plus proche, le statut d'urbanisation ainsi que des informations démographiques et économiques, définies à un niveau très fin. Conformément à la littérature existante (Olken, 2009; Yanagizawa-Drott, 2014; Miner, 2015), je constate que la distance du bureau de vote à l'antenne la plus proche et les caractéristiques du terrain sont les principales variables prédictives de la réception du signal. Par exemple, lorsque la distance à l'émetteur augmente de 1 km, la probabilité de recevoir le signal croît d'environ 60%.

J'effectue de nombreux tests d'exogénéité pour vérifier la validité de cette hypothèse d'identification, c'est-à-dire que l'accès au signal n'est pas significativement corrélé avec des covariables qui influencent également la participation politique, conditionnellement aux variables observables. Je constate que, en contrôlant par les variables prédictives du signal, les préférences politiques avant 2013, ou les appartenances ethniques, ne sont pas significativement corrélées avec la réception de la télévision en 2017. Je conduis des vérifications similaires en utilisant des données économiques et sociales – avant et au moment de l'élection – et les résultats ne suggèrent pas l'existence d'un biais dans l'attribution de la télévision, conditionnellement aux déterminants de la propagation.

Dans un second temps, j'étudie l'impact de l'accès à la télévision sur la participation lors de la première et la seconde élection. Les bureaux de vote "traités" sont ceux qui ont bénéficié de l'ouverture du signal entre 2013 et 2017, en contrôlant les déterminants précédemment identifiés. Je constate que l'exposition aux programmes télévisuels n'a pas influencé le taux de participation à la première élection (contexte électoral "normal"). Cependant, elle a eu un impact significatif sur le taux de participation lors de la seconde élection. Cet effet est différent au sein des deux principaux camps politiques (je sépare les bureaux de vote selon qu'ils aient voté en faveur du président élu lors de l'élection précédente de 2013). Lors de la seconde élection, le taux de participation a chuté de 4 points de pourcentage dans les bureaux de vote pro-opposition, mais il est 7% plus élevé dans les bureaux qui ont voté en faveur

du président élu en 2013. Une autre spécification du modèle suggère qu'une augmentation de 1% du score du président sortant ajoute environ 0,15 point de pourcentage à l'effet de base de la télévision. En d'autres termes, l'exposition à la télévision semble avoir dissuadé les électeurs de déposer un nouveau bulletin de vote lorsque l'opposition était en tête, alors qu'elle a plutôt amplifié le soutien au président sortant lorsque ce dernier faisait la course en tête.

La dernière partie du chapitre se pense sur les canaux possibles de l'influence de la télévision afin de mieux comprendre pourquoi le contexte électoral joue un tel rôle. J'étudie d'abord le contenu des émissions télévisées en construisant une base de données constituée de 86 189 reportages diffusés à la télévision. À l'aide de méthodes d'analyse textuelle, je documente le type de contenus diffusés au moment de l'élection. Le climat de tension et les émeutes en cours dans le pays ont été largement couverts par les chaînes de télévision au lendemain de l'annulation de la première élection. Par exemple, au cours des deux semaines précédant la nouvelle élection, la part des informations utilisant un vocabulaire spécifique à un lexique de "conflit" a augmenté de 10%. En outre, je souligne une baisse de la diversité des reportages diffusés, ce qui indique un changement par rapport aux choix de couverture habituels, et une plus grande concentration sur la crise politique en cours.

La force de la télévision en tant que média peut s'expliquer par l'exposition collective au medium dans les lieux publics. Au Kenya, environ 40% de la population possède aujourd'hui un téléviseur, alors que plus de la moitié de la population déclare s'informer par ce moyen au moins une fois par mois. Cet écart entre la possession de téléviseur et le nombre de téléspectateurs est révélateur de la dimension collective de l'exposition à la télévision, qui est souvent regardée dans des lieux publics. La fréquentation des cafés ou des restaurants est plus élevée dans les lieux couverts par le signal : je suggère sur cette base que la dimension collective de l'exposition à l'information peut se traduire par des mécanismes de coordination des comportements électoraux.

Ce chapitre s'inscrit ainsi dans une littérature plus large qui étudie les changements technologiques et l'impact de l'exposition aux médias. Il contribue au débat croissant relatif aux transformations de l'accès à l'information dans les pays en développement. Il suggère que les réactions individuelles à l'incertitude et aux contextes de crise peuvent être amplifiées par la couverture qu'en font les médias, notamment du fait de changements dans les lignes éditoriales, ce qui peut alors influencer le comportement des citoyens.

List of Figures

List of Tables